THE GREAT COURSES®

Topic
Science & Mathematics

Subtopic
General Science

# 12 Essential Scientific Concepts

## Course Guidebook

Professor Indre Viskontas
University of California, San Francisco;
San Francisco Conservatory of Music

**PUBLISHED BY:**

**THE GREAT COURSES**
**Corporate Headquarters**
**4840 Westfields Boulevard, Suite 500**
**Chantilly, Virginia 20151-2299**
**Phone: 1-800-832-2412**
**Fax: 703-378-3819**
**www.thegreatcourses.com**



Printed in the United States of America

## Indre Viskontas, Ph.D.

Cognitive Neuroscience Affiliate,
Memory and Aging Center,
University of California, San Francisco
Professor of Music,
San Francisco Conservatory of Music

Professor Indre Viskontas is a Cognitive Neuroscience Affiliate at the Memory and Aging Center at the University of California, San Francisco (UCSF), where she studies the emergence of creativity in patients with dementia. In addition, she is a member of the collegiate faculty at the San Francisco Conservatory of Music, where she is pioneering the application of neuroscience to musical training. She completed her B.Sc. in Psychology and French Literature at Trinity College in the University of Toronto, her M.M. (Master of Music) in Vocal Performance at the San Francisco Conservatory of Music, and her Ph.D. in Cognitive Neuroscience at the University of California, Los Angeles (UCLA).

Professor Viskontas's work is characterized by innovation and a focus on the big questions in neuroscience: How do brain cells code memory? How can the paradoxical facilitation of creativity emerge from a degenerating brain? What can neuroscience tell us about effective musical performance? Professor Viskontas has published more than 35 original papers and book chapters related to the neural basis of memory and creativity, including several in top scientific journals, such as the *Proceedings of the National Academy of Sciences*, *The Journal of Neuroscience*, *Current Opinion in Neurology*, and *Nature Clinical Practice*. Her work has been featured in Oliver Sacks's book *Musicophilia: Tales of Music and the Brain* and *DISCOVER Magazine*, among other publications. Her ongoing collaborations include projects with internationally acclaimed artist Deborah Aschheim, with whom she created art pieces highlighting the interplay among memory, creativity, and the brain, and a multimedia project exploring the interplay between musical ensembles and empathy, funded by a grant from the Germanacos Foundation.

Defying traditional career boundaries, Professor Viskontas spends much of her time performing as an opera singer—with favorite recent roles including Susanna and the Countess in Mozart's *Le Nozze di Figaro*, the title role in Gilbert and Sullivan's *Iolanthe*, Lazuli in Chabrier's *L'Étoile*, and Beth in Adamo's *Little Women*—with such companies as West Bay Opera, Opera on Tap, Lyric Theatre of San Jose, the Banff Summer Arts Festival, and others. She often works with composers and has created roles in three contemporary operas. A regular soloist with several chamber music groups, Professor Viskontas is the founder and director of Vocallective, a consortium of singers and instrumentalists dedicated to the art of vocal chamber music, as well as comanager of Opera on Tap: San Francisco, a chapter of the nationwide organization whose mission is to create a place for opera in popular culture by producing high-quality performances in nontraditional venues, such as art galleries, bars, and cafés.

Professor Viskontas's dissertation was recognized as the best of her class. She also has received numerous scholarships, fellowships, and awards, including a Julie Payette-NSERC Research Scholarship awarded to the top Canadian graduate students competing for the honor in the natural sciences and engineering, the Dr. Ursula Mandel Fellowship, a UCLA dissertation fellowship, the Charles E. and Sue K. Young Award for the top graduate students at UCLA, a McBean Family Foundation Fellowship during her postdoctoral work at UCSF, and the prestigious Laird Cermak Award given by the Memory Disorders Research Society. She also received the Distinguished Teaching Assistant Award at UCLA and served as the teaching assistant consultant in the Department of Psychology.

A passionate science communicator, Professor Viskontas made her television debut as cohost of *Miracle Detectives*, a documentary series with six hour-long episodes that aired on the Oprah Winfrey Network in 2011. She has appeared on *The Oprah Winfrey Show* and several major radio stations across the United States. She was a host of the podcast *Point of Inquiry* from 2012 to 2013 and currently cohosts *Inquiring Minds*, a popular science podcast produced in partnership with the Climate Desk. She is a sought-after public speaker, a fellow of the Committee for Skeptical Inquiry, and an editor of the journal *Neurocase*. ■

# Table of Contents

## INTRODUCTION



## LECTURE GUIDES

# Table of Contents

# Table of Contents

# 12 Essential Scientific Concepts

**Scope**:

What is the point of science? Arguably, the point is to understand the universe and our place in it and to use this knowledge to enrich our lives and reduce suffering. How has science changed the world? That question is what this course is about. The course will survey the scientific concepts that have changed our world and how we understand it, from quantum mechanics to emergence—from the smallest things to the unfathomable breadth of the universe.

But science is far from finished. If you think that science's most important discoveries are a bunch of settled facts and principles to memorize, you'll be amazed at how much we still don't know—at what is left to discover. After all, scientists devote their professional lives to exploring the unknown, not to confirming what we now take for granted as the truth. And many of the established concepts retain their mysterious essence. How else could you describe the following enigmas?

- Massive objects distort time, not just space.
- Light is made up of particles, but behaves like waves.
- Almost 10% of your DNA comes from viruses, and your DNA changes as you age.
- Your body contains 10 times more microbial cells than human cells.
- Much, if not most, of what you see with your eyes is made up by your brain.

The beauty of science is that with each question that is answered, many more questions are raised; each discovery helps us develop more refined queries about the world around us. Instead of reducing the wonders of the world to dull, boring facts, science illuminates and enhances them, by deepening our

appreciation of the world's complexity and the nature of human experience and fueling our sense of awe.

Each lecture in this course introduces and explains fundamental scientific concepts and puts them into our modern context, by describing both how they have been applied to make our lives better and what we still hope to learn. For example, understanding the nature of light led to the development of laser technology, and decoding human DNA has launched the era of personalized medicine—and both discoveries offer opportunities that we have only just begun to explore. The course starts by defining and exploring the very essence of life. Next, it investigates the origins of the diversity of life on our planet from the perspective of evolution and genetics. Then, it examines the world's most complicated and powerful organ: the modular and malleable brain. From there, the course probes the mysteries of magnetism and electricity, as well as the laws of thermodynamics and fluid mechanics that make so many engineering marvels possible. You will travel to the beginning of time to look at how the universe began and consider the key forces of nature. You will also go down the rabbit hole of quantum physics to understand the very nature of matter. Finally, you will reach the profound realization that out of the intricate organization that underlies biology, physics, math, and chemistry, some unexpected properties emerge that are greater than the sum of their parts. By the end of this course, you will be broadly proficient in scientific concepts that have broken the mold and changed humanity forever. ■

# The Miracle of Life

## Lecture 1

In this course, you will explore 12 of the most important concepts of modern science. Many of them, since their inception, have been relegated to separate disciplines—but, in fact, much like the way our brain cells work together, they are all closely linked, with each one playing an important role in our understanding of the world. In this lecture, you will learn that by breaking down life into its components, we can begin to grasp what makes it so special.

**Defining Life**

- What is the meaning of life? We seek answers everywhere, from the great works of literature to the most complicated mathematical theories of quantum physics. But despite our best efforts to find an answer, the meaning of life remains an open question in part because simply defining life is surprisingly difficult.

- The definition has to capture things that are very small and simple—like an amoeba or a bacterium—as well as the largest, most complex creatures, like humans and blue whales. It also has to encompass things that don't move on their own, like trees and strawberries.

- Instead of superficial features, maybe it makes sense to define life in terms of its properties—things that grow, reproduce, and consume energy. But many properties of living things also occur in inanimate objects: Mountains grow and recede, ocean currents move, and fire consumes energy. So what defines the border between the animate and the inanimate?

- Chemists classify matter into two kinds: organic and inorganic. All living things are made up of organic compounds, combinations of molecules containing only a handful of elements—carbon, nitrogen, sulphur, hydrogen, and oxygen being the main players. By definition, an organic compound is simply one whose molecules

contain carbon, an admittedly fairly arbitrary distinction. But it is curious that all living things contain carbon.

- Most contemporary scientists recognize that there is nothing fundamentally different between the elements that comprise animate and inanimate things. So, we come to the idea that life should be defined by its properties, and not the materials from which they are made. These properties can be divided into two categories: intrinsic and extrinsic. Intrinsic properties refer to internally driven actions—that is, from the thing itself—whereas extrinsic properties involve the interactions of the thing with its environment.

- The fundamental intrinsic properties include organization, growth, and reproduction. All things that we've ever identified as living are composed of one or more cells, which are the building blocks of life, and they are highly organized, even if they are single-celled organisms. Living things grow, by increasing in size rather than by simply swallowing other matter, and they are capable of reproduction—of producing new individual organisms, either asexually from a single parent or sexually from two parents.

- Living organisms also interact with their environments in characteristic ways, or have certain extrinsic properties: They transform energy by converting chemicals into cellular components, regulate their internal environments, adapt over time in response to their environments, and respond to stimuli or changes in their surroundings.

- All of these properties do not serve as an adequate definition of life, however, because some inanimate objects can show many of these properties, too. Chemistry isn't the whole story either, because we can build machines from biological parts.

- Life is a complicated concept. And when faced with complexity, scientists start by pulling apart the layers and considering one component at a time. Each of these methods—examining the properties of living things and analyzing their chemical makeup—

is useful in different circumstances as a way of understanding what makes life special.

- Science functions by finding converging lines of evidence to support ideas, following many different paths toward a comprehensive solution. And it's often useful to start with the parts and build up from there. So, to understand what life is, let's start with a closer look at the chemistry of living organisms.

**The Chemistry of Life: Water and Carbon**

- Despite the incredible diversity on Earth, the chemistry of life can be broken down into two critical items: water and carbon. Water is the molecule that supports all life, and carbon is an element that every molecule of life contains.

- All of the living organisms that we're familiar with are mostly water. About three-quarters of the Earth is covered by water. We now know that life on Earth started in the water and, for the first 3 billion years of evolution, stayed there. Most cells are

© Vladimir FLoyd/iStock/Thinkstock.

**Water is made up of two hydrogen atoms bonded to an oxygen atom ($H_2O$), and the hydrogen atoms of one water molecule are attracted to the oxygen atom in another to form bonds.**

surrounded by water, and even cells themselves are 70% to 95% water.

- Bathed in water, the stuff of life is laced with carbon, which is uniquely capable of forming large, complex, and diverse molecules. Other elements such as hydrogen, oxygen, nitrogen, sulfur, and phosphorus are also common in organic compounds, but it's carbon that takes the starring role.

- Water is a simple molecule made up of two hydrogen atoms bonded to an oxygen atom. Oxygen has more negatively charged electrons than hydrogen, and the bonds between the elements are single, covalent bonds—chemical bonds involving the sharing of electrons between atoms—that make the molecule slightly negative near the oxygen atom and slightly positive near the hydrogen atoms. This configuration makes the molecule polar, meaning that the overall charge is unevenly distributed across it.

- The life-giving properties of water stem from the fact that there are attractions between oppositely charged atoms of different water molecules. So, the hydrogen atoms of one water molecule are attracted to the oxygen atom in another. This attraction between molecules creates weak bonds that can form, break, and re-form very quickly and very often.

- This ability of water to organize and reorganize its molecules into a higher level of order provides the basis for four key emergent properties that explain its life-sustaining quality: cohesion, temperature moderation, solvency, and expansion.

- What do these qualities of water tell us about life? Can they help us set the boundary between the living and nonliving? Does water enable life to have some of the very properties that define it?

- The fact that water exists in all three states—solid, liquid, and gas—means that many different life-forms can take advantage of it, depending on their environments, and they can use this property of

water to regulate their own temperature. Water’s ability to dissolve many different compounds, while leaving other compounds unaffected, gives life-forms many options in terms of how to use water to perform basic functions. And the ability of water to defy gravity using its cohesive force enables living things to adapt their architecture in order to exploit this property.

- Perhaps water is one component that has made the diversity of life a possibility, and by being so versatile, it also gives life-forms a chance to adapt to surroundings and survive in the face of major changes. Water might be one key to the diversity, adaptability, resilience, and proliferation of life, all characteristics that seem to define the very essence of what it means to be alive.

- If water supports life, then carbon, arguably, is life. Carbon is a chemical element that has four valence electrons in its outer shell—electrons that it can share with other elements, making it possible for carbon to be the centerpiece in an almost infinite number of molecules. Carbon is ubiquitous in living organisms because it’s compatible with many other elements and because it can form complex chains with other carbon atoms.

- These chains are called carbon skeletons, and they form the backbones of most organic molecules. The wide variety of shapes in which the chains can form underlies the diversity of organic compounds and, therefore, of life itself.

- Organic compounds are classified into four general categories: amino acids, which make up proteins; nucleic acids, the building blocks of DNA (deoxyribonucleic acid), our genetic code; carbohydrates, such as sugars, starches, and cellulose; and lipids, which are fats and hormones.

- These four different categories represent both the incredible diversity and the adaptability of carbon-based compounds. With the same building blocks of elements, the entire catalogue of living things can be created.

**Proteins: The Worker Bees**

- Proteins are the organic compounds that enable cells to survive, reproduce, adapt, consume energy, fight enemies, self-regulate, and perform every other function that seems to characterize life. Each cell in our bodies works hard to design, manufacture, store, and transport proteins. Nearly everything our cells do involves proteins.

- Proteins are made of amino acids. An amino acid is a molecular compound that has a distinctive chemical structure: At its very center is a carbon atom, which has four possible bonds that can be filled with other atoms or compounds. In amino acids, three of these bonds are always the same. What distinguishes one amino acid from another is what's attached to the fourth bond.

- Despite the wide array of different kinds of proteins, all of them are made of some combination of the same 20 amino acids. A string of amino acids is called a peptide, and strings of peptides are called polypeptides. A protein is a molecule made of one or more polypeptides, or strings of amino acids, that's folded into a particular three-dimensional structure, which is very important because the successful function of a protein depends on its shape.

- Within our bodies, we house tens of thousands of different proteins with different shapes and functions. Almost every protein works by recognizing and binding to some other molecule, so if the shape is off, it can't function.

- Many proteins are tasked with the job of speeding up chemical reactions—as catalysts. These proteins are called enzymes, and they can perform their duties over and over again, making them the true worker bees in the cell.

- Other proteins help our cells store important molecules for later use, or transport them around the body, or aid in the communication between cells, or coordinate activity between different organs, or provide structural support. Without proteins, life would be impossible.

## Suggested Reading

McGee, *Bioethics for Beginners*.

Reece, Urry, Cain, Wasserman, Minorsky, and Jackson, *Campbell Biology*, chap. 2–5, p. 30–93.

## Questions to Consider

1. How would you define life?

2. Where is the boundary between “real” and “artificial” life?

# The Organization of Life

## Lecture 2

Life defies simplicity, in any form. And this is one property of it that we have yet to address in our search for an adequate definition of life: its complexity—the organization that characterizes every life-form, from amoebas to antelopes. Life is highly organized. Organization equals capability, and capability equals survival. And whatever else life may be for, every example of it that we know seeks to perpetuate itself.

**Prokaryotic versus Eukaryotic Cells**

- Cells are the building blocks of living organisms; they are the smallest, organized, self-contained living systems. They can be entire organisms by themselves or only one of billions of cells making up a complex human being.

- We all came from the same starting point and, despite the long and convoluted path of evolution, have retained many features of our tiny one-celled ancestors. By some measures, 5% to 7% of our DNA is the same as that in bacteria. We truly are connected to all other living things on Earth.

- Cells are tiny bags full of organic compounds, but just as those compounds are intricately organized chains of carbon and its atomic buddies, cells have their own structure and key bits.

- Broadly, there are two types of cells in living things: prokaryotes and eukaryotes. Prokaryotic cells carry their genetic material in a region that is not enclosed by a membrane like the nucleus in a eukaryotic cell. Eukaryotic cells protect their genetic material with a special double-membrane wall that surrounds the nucleus of the cell.

- Prokaryotes are mostly bacteria. Why wouldn't bacteria want to protect their DNA? Prokaryotes are thought to be the first living cells to evolve, so they're phylogenetically older than eukaryotes.

The membrane surrounding the nucleus evolved later. And they've also done pretty well without nuclei.

- They are very well adapted to harsh environments. We don't fully understand why prokaryotes are so adaptable, but their hardiness and adaptability have made them the most abundant life-form on Earth.

- Bacterial cells are typically much smaller than eukaryotic cells and have a variety of shapes. They are highly successful in part because they can reproduce rapidly when the circumstances are right. Under the best conditions, some species can replicate a whole new generation in less than 20 minutes. This rapid rate of replication gives bacteria the opportunity to profit quickly from random mutations.

- With short generation times and large populations, even rare new mutations can give the species an advantage when it comes to surviving new environments, because they have more opportunities to develop mutations that might protect them from the elements. This is one of the ways in which bacteria can quickly develop a resistance to antibiotic medicines.

- Another quality that enables antibiotic resistance is the unique mechanisms by which prokaryotes transfer genetic material, or DNA: They can take up genetic material from their surroundings and incorporate it into their own code—even when the new genes come from a different species.

**A Tour of the Cell**

- Despite the fact that 1% to 3% of our body mass is bacteria, the rest of our bodies consists of eukaryotic animal cells. Remember that all cells need to sustain and regulate themselves, reproduce, and help the organism grow. All of these functions require cells to be highly organized.

- The cell's membrane, or cell wall, is made up of lipids—so it's water fearing, or hydrophobic. This is so that the contents of the cell don't leak out. However, if the entire membrane were hydrophobic,

then the outside of the cell would repel water that carries nutrients to the cell, like oxygen and glucose. The cell needs these nutrients to survive.

- Structurally, lipids resemble jellyfish—they have a spherical head attached to some chains that bring to mind long tentacles. The spherical head is hydrophilic, or water loving. The tentacles are hydrophobic. The cell membrane has two layers of lipids organized so that the spherical heads are on the outside and the tentacles meet on the inside.

- This organization allows the spherical heads that are water loving to be in contact with the water. Scattered throughout the membrane are proteins that act like gates: They help some desirable molecules get across the cell wall and keep the unwanted ones out.

- Once we get past the cell wall, the next most obvious part of the cell is the nucleus, which also contains a membrane in eukaryotes. Just like the outer cell wall, the nuclear membrane has a lipid structure, but for added security, the membrane is doubled—that is, there are two membranes, each with a lipid bilayer. The wall also has pores, or holes, that facilitate the passage of certain large proteins and macromolecules.

- The main job of the nucleus is to house and protect the cell's genetic information, which is stored in the form of DNA. Proteins are what make a cell function and display the properties of life, and to make proteins, the cell needs to use the DNA blueprint to manufacture and assemble these complex molecules.

- Inside the nucleus, the all-important job of using the DNA code to generate messenger RNA (mRNA) occurs—the first step in the creation of new proteins. Each gene along a DNA molecule permits the synthesis of a specific mRNA, which then interacts with the cell's protein factories to direct the assembly of proteins and their parts.

- In the nucleus, the DNA creates mRNA, which is then sent out to the protein factories via the pores in the membrane surrounding the nucleus. These protein factories are called ribosomes, and they come in two types: free and bound.

- Free ribosomes float around in the cell while bound ribosomes are attached to the endoplasmic reticulum, a network of sacs and tubes that serve as the industrial zone for the cell's various large manufacturing projects, such as the synthesis of new membranes or protein gates.

- All of these different components of the cell underscore the very fundamental characteristics of life: sustenance, reproduction, regulation, and even adaptation. There are three more key parts to the cell: the golgi apparatus, the cell's storage and shipping warehouse; the mitochondria, the energy generator; and the lysosomes, the waste management system.

- In organisms as complex as humans, there are many different types of cells—from neurons to pancreatic cells to blood cells—and each one has additions or subtractions that support its specific function. The ways that those different cell types interact took hundreds of millions of years to evolve. But the cell parts discussed are common to most, if not all, cells and serve as the building blocks for our bodies.

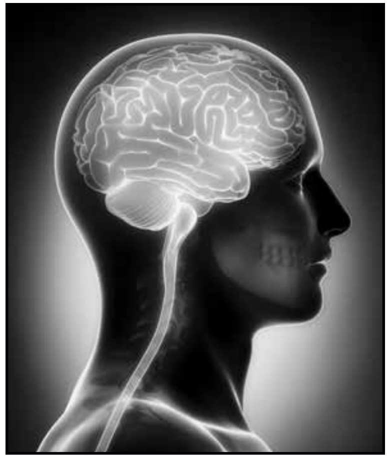
**Viruses: Living or Nonliving?**

- There is some debate as to whether a virus should be considered a living or a nonliving thing. A virus cannot replicate or power itself; it has no metabolism or ability to generate energy from its environment. It depends on host cells to make new proteins.

- However, a virus is made up of the molecules of life, DNA, so it can replicate only with the help of a host and, therefore, evolve. These features make classifying viruses as living or nonliving difficult.

## The World's Most Complicated Organ: The Brain

- The complexity of single-celled organisms is remarkable, but in some ways, even more remarkable is the fact that billions of seemingly self-sufficient cells can work together to form complex beings like humans. If we really want to understand why one of the fundamental properties of life is organization, we need only to peek inside the most complex thing we've discovered so far in our universe: our brain.


© Janulla/iStock/Thinkstock.

**The human brain, which is made up of about 100 billion cells, is the most complicated organ.**

- Our brain is made up of about 100 billion cells, or neurons. Each of these neurons is connected to other neurons, and information flows from one cell to the others. Some neurons get their information from as many as 10,000 other cells and send it on to thousands more.

- The signals that are sent between neurons are binary, just like the code that our computers use. They either encourage the downstream neuron to send its own signal or to keep quiet. However, if we look more closely at the junction between one neuron and the next, the story quickly becomes far more complicated than that.

- The space that separates one neuron from another is called a synapse. More specifically, a synapse is the spot where one neuron sends its information and another receives it—like a dock in a shipyard.

- Ships come in with containers full of different products. Many of those containers are accepted at the dock, and they release their contents into trucks or other vehicles that then transport them to other locations. Some containers can get lost during the

transportation; some are rejected by customs. Some ships don't even drop off cargo; they just deliver instructions for the customs officials to use.

- At the synapse, many neurons dock their axons—the part of the cell that sends out electrical and chemical signals. The receiving neuron, or the destination, also has many docks, or dendrites, to receive the signals.

- Scattered along the dendrites are customs gates, which analyze the structure of the molecules that are released by the axons of the upstream neurons. Some of those molecules are accepted, and the signal gets propagated down the dock and into the core of the receiving cell. Some are rejected and never make it past the synapse. And some are lost during the journey; they never make it to the dock.

- All of these different molecules and their interactions have an effect on the functioning of the cell and its downstream connections. Some molecules are involved in making memories, and they change the way that the neurons get connected. Some turn on or off other functions, like those that cause the cell to fire off a signal or stop it from firing.

- Then, if we move a level up in terms of the organization of the brain, we find that there is a pattern to the electrical signals that neurons generate. Certainly, these signals are affected by neurotransmitters, the molecules that travel across the synapse, but there is also a code in the firing patterns that can be just as important in terms of making memories or relieving pain, for example.

- Cells send signals by altering the rates at which they fire, or by firing in bursts or in synchrony with one another. Each of these codes can be related to different brain functions and malfunctions.

- The brain's intricate organization can also be seen in the neuroanatomy of different brain regions and the connections

between them. For example, the hippocampus's complex neural architecture is responsible for creating memories.

## Suggested Reading

Reece, Urry, Cain, Wasserman, Minorsky, and Jackson, *Campbell Biology*, chap. 6, p. 94–124.

Zimmer, *A Planet of Viruses*.

## Questions to Consider

**1.** Why do viruses represent the border between the living and the nonliving?

**2.** How are multicellular organisms different from unicellular ones, from the perspective of the definition of life? Is a single cell taken from a living organism alive? Are some cells more "alive" than others (i.e., is an egg cell more alive than a skin cell?)?

# Evolution—The Tireless Tinkerer

## Lecture 3

In the words of Charles Darwin, evolution is "descent with modification." That is, today's species are modified descendants of ancestors that inhabited the world in days past. The pattern of evolution is seen in data from a myriad of disciplines, from geology and physics to biology and chemistry. These disciplines provide us with the facts of evolution—observations of how the natural world has changed. The process of evolution refers to the mechanisms that cause these patterns, and if anything remains open to modification, it's the mechanisms by which evolution occurs that remain theoretical.

**The Facts of Evolution**

- Until the mid-1800s, most people believed that every animal on Earth was created as we see it today and that species do not change—that the uniqueness and diversity of species was a testament to the power and creativity of God. Charles Darwin changed all of that.

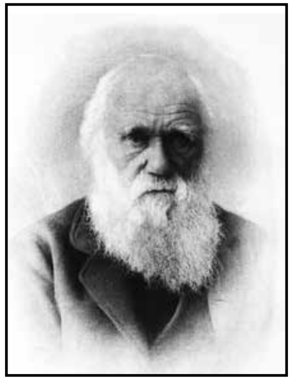
- However, like that of every scientist, Darwin's work built upon the observations and missteps of others; science progresses because scientists constantly question and check the work of those who go before them.

- Darwin's five-year journey aboard the HMS *Beagle* forever changed our understanding of biology. But when he first set foot on the ship, evolution still seemed to him a far-fetched notion, supported by only a fringe group of radical thinkers.

- We often think of Darwin as a pioneering adventurer sailing the world's seas for the love of science. But seasickness made him a terrible sailor and plagued him throughout his voyage. And he even made some rookie mistakes when it came to collecting specimens. And his ideas, ultimately, weren't completely original; many of them have roots in the writings of scientists who came before him.

- But what made his treatise, *On the Origin of Species*, so influential is the mountain of evidence that Darwin supplied to back up his thoughts. He found converging evidence from different scientific domains—such as geology, biology, and geography—but he also pointed out where data were still missing and where scientists should look to disprove his ideas.

- There remained many mysteries and many holes in the fossil record that could show how one species gave rise to another. Since Darwin's observations, many of these holes have been filled with new evidence, and as a testament to the veracity of his theory, his ideas have withstood the test of time and the emergence of new data intact.





**Charles Darwin (1809–1882) proposed the theory of evolution by natural selection.**

- The Galapagos Islands—off the west coast of South America, in the Pacific Ocean—provided the "aha" moment that led Darwin to formulate his masterpiece. Ironically, he didn't realize the significance of his observations on the islands until long after he returned home. But what captured his imagination was the diversity and uniqueness of the plants and animals that he saw in the Galapagos.

- Darwin collected specimens of birds from the various islands and eventually sent them to John Gould, a British ornithologist. Darwin had observed different beaks belonging to different birds and had assumed that they were all different species of finches, wrens, and blackbirds.

- However, many years later, Darwin went to hear Gould talk about his birds and was astonished to hear him say that they were, in fact, all finches. They simply had different beak shapes so that they could eat different foods.

- Darwin realized that not only had he failed to notice this diversity within the same category of bird, but he had also been careless in his note-taking and hadn't written down the island from which each bird had been collected. Fortunately, unlike Darwin, the crew from the HMS *Beagle* proved to be much more meticulous; many had noted the specific islands which housed each bird.

- From these notes, Darwin found out that different types of finches lived on different islands, and he was puzzled by the diversity that characterized these relatively close and similar islands. It turns out that the abundance of particular types of food is one of the properties that distinguishes one island from another. And the finches that populate a particular island have beaks that are adapted specifically to the type of food that is most easily consumed in that environment.

- The realization led to the following question: If finches in the Galapagos descended from one ancestral species, and their diversity resulted from adaptations to specific environments, could the same process explain the evolution of human beings?

**Divergent versus Convergent Evolution**

- The Galapagos finches demonstrated the diversity of variations between closely related species, but Darwin observed another branch of evidence supporting the theory of evolution that came not from differences but from similarities. Evolution, according to Darwin, is the descent with modification from an ancestral organism as its descendants face slightly different environmental conditions and adapt to them accordingly.

- Depending on the environmental conditions (including the other organisms that are present in a particular habitat), anatomical structures present in a common ancestor can be harnessed for different purposes by different species. These body parts are called homologous structures, and the similarity present in different species descending from the same ancestor is called homology.

- One example of homology is found in development, when every vertebrate embryo starts with a tail and throat pouches, which turn into gills in fishes and ears and throats in humans. Some homologous structures continue to characterize a species even when they are no longer useful.

- When this happens, the features are called vestigial structures—vestiges of an animal's ancestor. An example of a vestigial structure is the pelvis and leg bones found in some species of snakes.

- But not all similar structures across species indicate descent from the same ancestor; some body parts result from convergent, as opposed to divergent, evolution. Divergent evolution describes how the same ancestor can produce different species with different traits. Convergent evolution describes the independent evolution of analogous, or similar, structures in different species resulting from adaptations to similar environments. So, in convergent evolution, similar traits come from different ancestors.

**Evidence for Evolution**

- Homologous and analogous structures are one type of evidence for the process of evolution. A second domain in which evidence is found is the fossil record. We look to the fossil record to find out what types of species roamed the Earth at different points in time.

- In Darwin's day, the fossils that paleontologists had discovered only went back so far in time, and the animals that left their mark were found to be large and quite distinct from those currently alive, but no less complex. So, many scientists of the $19^{th}$ century interpreted this evidence as showing that species might go extinct, but they do not evolve from each other as descent with modification suggests.

- Over time, however, as paleontology progressed and more and more fossils were uncovered, the data provided more support for evolution than for the idea that species are static. Today, there is a vast amount of evidence showing that past species not only differed

from current species, but also were progressively less complex the farther back we go in time.

- There is also a vast amount of evidence showing that certain early organisms gave rise to many different versions of more complex forms, with the biodiversity across the ages resembling a tree with more and more branches rather than a series of lines ending in oblivion.

- Critics of the theory of evolution sometimes point to holes in the fossil record as evidence that the theory has yet to be proved. Certainly, the fossil record is not a complete catalogue of every organism that ever roamed the Earth, but it is a remarkably rich source of evidence supporting evolution.

- A third type of evidence for evolution comes from the geographic distribution of species, or biogeography. By understanding continental drift and the movement of landmasses over time, we can predict where fossils of particular organisms should be found and how one species gave rise to others.

- For example, around 250 million years ago, there was only one supercontinent on Earth, called Pangaea. Over time, as different parts splintered off, new environments were created that forced species to adapt to different conditions if they were to survive. By charting the emergence of these adaptations and comparing them to the geological changes resulting from continental drift, we begin to understand why some species that currently live on different continents seem to share the same ancestor—or not.

**The Theory of Natural Selection**

- Darwin observed that there is individual variation in a species in terms of hereditary traits and that species produce more individuals than their environments can support. As a result, many of these offspring fail to survive and/or reproduce, creating competition for survival among individuals in a population.

- From these observations, he inferred that those individuals whose inherited traits give them a better chance of survival in their environment tend to leave more offspring than others—and that, over time, this ability of certain individuals to survive and reproduce shifts the proportion of the desirable traits in a population, leading to the gradual accumulation of changes adapted to the environment.

- Natural selection is the process by which individuals with certain inherited traits produce more viable offspring, which leads to a better match between the species and its environment over time. Of course, if the environment changes due to some catastrophic event (such as a flood or global warming), natural selection may result in new adaptations or even the rise of a new species—or the species can go extinct.

- Darwin coined the term "natural selection" to distinguish it from "artificial selection," which farmers and breeders had already been practicing for hundreds of years by the time he proposed his theory.

- Whereas farmers and animal breeders artificially chose animals with desired traits and bred them together to change the frequency of those traits within their population, natural selection does not work with such a determination in mind. Today, the term "artificial selection" has been replaced by "selective breeding."

- Natural selection works on a population, not on an individual. That is, it's a population that evolves, not a single organism. An individual might have a hereditary trait that makes him or her more likely to survive, but it's only across generations that we can see the effects of natural selection—only when we compare the genetic makeup of the individual's generation with that of his or her children's. In addition, natural selection can only work on heritable traits that differ among individuals; if all organisms in a species share the same trait, that trait cannot change.

## Suggested Reading

Bergstrom and Dugatkin, *Evolution*.

Zimmer, *Evolution*.

## Questions to Consider

**1.** What was so frightening to Darwin about the idea that humans evolved from monkeys? Does evolution diminish the precious nature of life?

**2.** Why do we say that Mother Nature is a tinkerer, not a designer? What is the evidence for this view?

# Other Mechanisms of Evolution

## Lecture 4

Natural selection isn't the only way by which a species can change over time; there are other mechanisms of evolution. In this lecture, you will examine some of the forces that drive evolutionary change—from genetic drift and gene flow to natural selection—which leads to consistent adaptations to specific environments. You will also learn how to tell if a population is evolving, and then you will learn how that information can be used to infer the mechanism that is driving the changes.

### Genetic Variation

- Evolution wouldn't happen if every individual in a species were exactly the same; individual variability is a prerequisite for natural selection to take effect. And this variability needs to be tied to something that gets passed down from one generation to the next: It has to be hereditary.

- In Darwin's day, scientific knowledge of genetics was only just beginning, and Darwin himself didn't know about genes. But now we know quite a bit. And we can see natural selection in action even at the smallest scale—what is known as microevolution, which describes the changes in the frequency of specific variants of genes in a population over the course of several generations.

- Genetic variation refers to the differences between individuals that are tied to their DNA, or genetic code. Not all of our differences are hereditary. For example, body shape does have a genetic component, but life choices can have a huge effect on how someone looks.

- Our genes are encoded on a set of 23 unique chromosomes, strands of DNA that contain many genes, and we have two copies of each chromosome—one from our mother and one from our father, for a total of 46. So, we inherit two copies of each gene, one from each


**Deoxyribonucleic acid (DNA) is made up of a double helix that is held together by hydrogen bonds.**

of our parents. These genes are the blueprints for all of the proteins that we need for our cells to function.

- These two copies are called alleles, and they can either be the same (homozygous) or different (heterozygous). When we're looking for evidence of microevolution, we look for changes in the frequency of one allele or another—that is, over generations, does one specific type of gene become more frequent in the population?

- The interesting thing about microevolution is that natural selection is not the only mechanism by which the allele frequency in a population can change, although it happens to be the one mechanism that regularly leads to an organism's adaptation to its environment. Two other mechanisms that can change allele frequency are genetic drift, which is caused by random sampling, or the effect of chance; and gene flow, or the transfer of alleles between populations.

- It's not a given that a population can evolve; there are certain conditions, such as genetic variability, that have to be in place. We have a way of assessing the likelihood of whether a population is evolving: the Hardy-Weinberg law, named after G. H. Hardy and Wilhelm Weinberg, who first demonstrated it mathematically in the early 20th century.

**The Hardy-Weinberg Law**

- The Hardy-Weinberg law determines the allele frequency in a population that is *not* evolving. That is, allele frequency should remain constant from generation to generation when the population is not affected by natural selection or other causes of allele frequency shifts, such as genetic drift and gene flow.

- Once we have the predicted frequency from the Hardy-Weinberg law, we compare it to what we see in the actual population. If they are different, then we can say that the population is in the process of evolving.

- When the principle accurately predicts allele frequency from one generation to the next, we assume that the population is in equilibrium. Departure from equilibrium occurs when one of the following five conditions is *not* met. That is, these conditions describe a population whose genes aren't changing.
  1. There can be no mutations.

  2. The population should engage in random mating.

  3. There cannot be natural selection.

  4. The population should be very large.

  5. There can be no gene flow.

- If these conditions of the Hardy-Weinberg law are met, it means that the population is isolated—either geographically or culturally.

- The Hardy-Weinberg equation gives us a starting point to compare our population of interest against. By evaluating which of the conditions are not met and why, we can begin to speculate and eventually prove which mechanisms are causing the change in allele frequency.

**Other Mechanisms of Evolution**

- Natural selection—or, very roughly, the survival (and reproduction) of the fittest—is one mechanism by which evolution takes place. Using the Hardy-Weinberg assumptions, we can figure out what other forces cause evolution if we examine which of the assumptions the particular population we're studying has violated.

- Let's start with the first condition—that mutations do not occur that could change the alleles. We know that's not true for most populations, because mutations do occur frequently and spontaneously, usually as a result of mistakes that happen during DNA replication.

- If an accidental mutation happens early in pregnancy and the embryo never develops, that mutation has no chance of being passed on to the next generation. But there are also random mutations that occur that are less devastating and don't impede reproduction.

- You might think that such mutations would have the largest effect on the evolution of a population; a direct change in the genes, after all, seems like the most efficient way to change alleles. But to understand the potential of mutations to alter populations, we first must consider the vast number of genes that form the code of our lives.

- The human genome contains the blueprints for about 20,000 to 25,000 genes on our 23 pairs of chromosomes. That seems like a lot of genes when you consider that these spontaneous mutations occur at random. Even if there are many mutations in a population, there are 20,000 genes to choose from, so the mutation rate for any given gene is pretty small.

- And for this random mutation to have a cumulative effect that's visible across the population, the rate of mutation would have to be very significant indeed—so high, in fact, that it's unlikely that the species could still function well.

- These random mutations are in fact an important prerequisite for evolution to be possible: Although mutations aren't a force of evolution in and of themselves, they set the stage upon which other evolutionary mechanisms can act. Because random mutations are, by definition, random, they don't have a big influence on the rate of alleles in a large population.

- The second assumption of the Hardy-Weinberg law that is sometimes broken is that mating is random. Random mating ensures that any mutations that do arise are dispersed among the population, rather than concentrated in one group. The negative consequences of nonrandom mating can be devastating.

- The perils of inbreeding, for example, are not uncommon among humans—just look at the rates of hemophilia, psychosis, and other genetic diseases among members of royal families. These diseases are more common when relatives mate with each other because the chances of a child inheriting a bad copy of a gene are greater when there is less variability in the genes of his or her parents. So, familial diseases become more frequent.

- What if the population that we're studying isn't very large? Violating this assumption of the Hardy-Weinberg law can lead to genetic drift: when allele frequency shifts because of random selection effects. In small populations, genetic drift can have a big influence, decreasing genetic variability and therefore making the population vulnerable to diseases or extinction very quickly.

- The last assumption of the Hardy-Weinberg law is that genes don't flow into and out of a population—that is, individuals from one group don't mate with those from another population. We're not

talking about inbreeding here; in both cases, the populations are very large.

- For a population to be in equilibrium, with allele frequencies remaining constant, which is what the Hardy-Weinberg law is designed to detect, alleles can't be flowing into and out of a population—as would be the case if individuals were flowing into and out of it and mating with members of different populations.

- Think about human immigration: There are all kinds of people moving between countries, and they are not being sorted out in terms of their alleles. So, like random mutations, the effects of gene flow on allele frequency are fairly low. In fact, they also cause the opposite effect; they can reduce genetic differences between populations, bringing the two groups closer together genetically.

**Natural Selection in Action**

- Natural selection provides an explanation for the diversity of Mother Nature and the observation that species are precisely adapted to their environments. It's important to understand that evolution by natural selection is *not* random: Mutations are random, but sorting by natural selection favors some alleles over others, simply because the realities of any environment favor the attributes of some organisms over those of others.

- Evolution by natural selection is sometimes described as the "survival of the fittest." It's important to remember that we're talking about fitness as it relates to the ability to produce more viable offspring. So, competition that occurs after the age of reproduction has no effect on evolution—unless, of course, it affects the likelihood that the offspring themselves will reproduce.

- We are most interested in relative fitness, which is how much an individual contributes to the gene pool compared with other individuals. Greater fitness might be demonstrated in individual battles between gorillas to attract a mate or in the ability of a moth to blend into a tree, thereby avoiding being eaten long enough to

have babies—or in the skill of a barnacle in obtaining a greater proportion of available food, thereby depriving its neighbors of adequate nutrition and preventing them from reproducing.

- One of Darwin's most original contributions was the description of sexual selection and its effects on a species. Sexual selection refers to the process by which some inherited traits make an individual more likely to find and secure mates. Sexual selection is a subtype of natural selection. A consequence of sexual selection is sexual dimorphism—when the two sexes in a species develop very different characteristics. An example is peacocks and peahens.

- Sexual selection seems to be a particularly potent form of natural selection because it acts directly on reproduction. The ability to survive long enough to reproduce is essential, but if you can't find a mate, you definitely won't be passing on your genes.

## Suggested Reading

Darwin, *On the Origin of Species*.

McKenna, *Superbug*.

## Questions to Consider

**1.** Why does natural selection lead to adaptations to the environment, but other mechanisms of evolution, such as genetic drift or gene flow, do not?

**2.** If we begin to control reproduction by "designing" babies and artificially selecting certain traits in humans and animals, how does this activity affect evolution? Are we still evolving then?

# DNA and Heritability

## Lecture 5

Some idea about heritability, the idea that traits are somehow transferred from one generation to the next, has existed for centuries. How this happens—that is, what the biological mechanism of inheritance might be—was unknown until in the middle of the last century. That mechanism is now the expansive field of genetics. In this lecture, you will explore several fundamental topics in genetics, including what a gene is, how a gene stores information, and how that information is used by the body to perform various functions.

**Mendelian Inheritance**

- It was through the pioneering work of Gregor Mendel that the field of genetics was born. His passion for crossbreeding pea plants revolutionized biology and medicine forever. Mendel is known for his many seminal experiments on the patterns of inheritance when different varieties of pea plants were selectively crossbred. Mendel's work would ultimately support Darwin's theory of evolution.

- It was well known in Mendel's time that creatures large and small pass on traits to their offspring. The dominant theory, however, was one of "blending"—that the characteristics of the parents are somehow combined, the result being an offspring having traits that were related but not direct replications of parental traits (rather, a mixing of the two).

- Like people, many plants use sexual reproduction, involving two parents, to create offspring, and pea plants are among them—it's called cross-fertilization—although pea plants can self-fertilize as well.

- Mendel was interested in how traits were passed from one generation to the next. He often worked with two strains of pea

plants, one with white flowers and one with purple flowers. He selectively interbred these plants to see how the offspring of his crossings turned out.

- According to the blending theory, mating white and purple flowers should produce offspring with a flower color approaching purple or pink, something not quite white and not quite purple. Instead, Mendel found that the offspring plants had distinctly white or purple flowers—not a blending of the two. The traits that were inherited were discrete, not mixed.

- But he also observed that the numbers of white or purple flowering plants in the next generation followed a specific pattern, and this discovery is what led to a paradigm shift in our thinking about inheritance.

- From his simple experiments, Mendel surmised two generalizations about inheritance that are now known as Mendel's laws of inheritance. These are the law of segregation and the law of independent assortment.

- The law of segregation states that, for every characteristic, such as eye color or hair color, each parent possesses two possible versions of that characteristic and that during the reproductive process, these two traits, known as alleles, are separated (that is, segregated), with only one version passing on to the offspring.

- Mendel conceived of these separate versions as dominant and recessive traits. Going back to the pea plants, purple was the dominant flower color while white was the recessive color, because when one white plant and



**Gregor Mendel's crossbreeding of pea plants led to his theory of inheritance.**

one purple plant were crossed, the offspring were always purple. Why do white flowers show up again in the third generation? That's where the second law that Mendel offered fits in.

- The law of independent assortment states that alleles of multiple traits segregate and are shuffled independently of one another. If you have two traits that get passed down, the alleles in each offspring are independent—that is, having one allele of trait A doesn't have any effect on which allele of trait B the offspring is likely to inherit.

- Mendel derived this law from experiments in which he tracked two traits—such as flower color, pea size, or pea shape—across pea plant generations. In essence, he found that the inheritance of these separate traits also followed patterns with three-to-one ratios, reflecting their own laws of segregation.

- What Mendel did not know—could not know—during his time was the biological mechanism underlying the inheritance patterns he observed. He inferred the existence of a "heritable factor," but we know today that the units of inheritance are genes, with two alleles, one from each parent.

**Genes and DNA**

- A gene is a distinct chain of DNA—that's a chemical compound called deoxyribose nucleic acid—that contains the biochemical instructions for the production of a protein. Genes and a lot of so-called junk DNA (called introns) are coiled tightly into little bundles called chromosomes. Junk DNA doesn't provide any code for proteins, the usual function of DNA, but it likely has some other biological function related to genetics. We just don't know what that is yet.

- Humans have two sets of 23 chromosomes, 46 in total, in the nuclei of our cells. One set of chromosomes comes from each of our parents, so every gene has two alleles: one from the mother and one from the father.

- You can think of DNA as a simple language—a code, like Morse code—that can be read by a cell's machinery to produce new cellular components called proteins. If cells are the building blocks of our bodies, proteins are what keep them functioning. They are the agents that are responsible for all of the actions of our cells.

- DNA utilizes four nucleobases (or bases, for short), which are four specific compounds that each contain nitrogen. These bases can be thought of like letters in the alphabet, such that different combinations of them form different genes, just as different combinations of the same 26 letters of the alphabet make up many different words. In fact, these four letters, or bases, make up the entire alphabet in the genetic code.

- The four bases are adenine (A), guanine (G), cytosine (C), and thymine (T), each made up of different combinations of nitrogen, carbon, oxygen, and hydrogen. Their different combinations of elements give them different shapes.

- DNA exists as a so-called double helix, which looks like a twisting ladder. The twisting DNA ladder, the double helix, is actually two separate strands of DNA bound together, like a zipped-up zipper.

- The rungs of DNA's twisted ladder are the bases—A, G, C, and T. The bases have natural pairs, or natural preferred partners to which they will bind. A and T bind together, while G and C prefer each other. It is the pattern of these bases that makes up our genetic code.

- Genes are an instruction manual for the production of proteins. At the molecular level, proteins are just precisely folded and decorated polypeptides—that is, amino acids strung together. Amino acids are small molecules that are the building blocks of proteins, and Mother Nature uses 22 of them in the construction of a wide variety of proteins.

**Turning Genes into Proteins**

- The process of turning the DNA in genes into proteins involves two important steps called transcription and translation and a few biochemical players to help get the job done.

- In essence, transcription and translation involve copying the information in DNA, putting it into a portable form, and transporting it out of the nucleus to other parts of the cell, where the cell's manufacturing machinery can construct the new protein.

- The process of transcription copies the DNA code into a new message called ribonucleic acid (RNA). DNA and RNA are both nucleic acids, and almost the same bases are the letters of their alphabets. Like DNA, RNA also has a four-base alphabet with which it stores information, but there is one small variation: RNA uses uracil (U) instead of thymine as the binding partner for adenine.

- A key player in DNA transcription is an enzyme called RNA polymerase that performs the transcription process. RNA polymerase binds to the DNA and acts to separate the DNA's bound base pairs, effectively unzipping the double helix. As it does so, moving across the DNA strands like the slider body of a zipper, the RNA polymerase reads the DNA sequence and creates a complementary, matching sequence of RNA.

- The double helix re-forms behind the RNA polymerase, and the new RNA sequence is released, a fresh Xerox copy of the DNA from which it was derived. The new RNA sequence now contains the instructions for assembling the peptides that will make up a new protein. However, remember that a lot of the information in our DNA is junk—or, at least, its purpose is not entirely clear, but it's not involved in creating new proteins.

- One intermediary step that occurs after RNA synthesis is that the unneeded bits of RNA, copied from the pieces of junk DNA, are cut out, leaving just the RNA needed to create the new protein. So,

now we've got our messenger RNA, and we use that to construct a new protein—a process called translation, which is how the nucleic acid code of the RNA is deciphered, or translated, into a polypeptide sequence.

- There are 22 amino acids that are used to construct the peptide sequences of proteins, and there are four bases used in RNA, so clearly there cannot be a one-to-one correspondence between the identity of the base and the corresponding amino acids. It turns out that triplets of nucleotides correspond to specific amino acids, like little three-letter words. For example, the triplet UGU corresponds to cysteine, while GCU stands for alanine.

- These triplets are known as codons, and there are 61 that code for amino acids. There are three additional codons that code for the instructions "start" and "stop," similar to a telegraph message.

- The key player in messenger RNA translation is the enzyme complex called a ribosome. Similar to the way RNA polymerase interacts with DNA, the ribosome binds to the RNA and moves across the sequence, reading the codons in turn and catalyzing reactions that sequester and add amino acids, found in the soupy milieu of the cell, into an elongated chain. The creation of the elongated chain ends when the ribosome encounters a "stop" codon.

- The product of this process is the polypeptide that will become part of a protein or a protein in and of itself. The entire process of translation occurs very quickly, in a fraction of a second. In addition, a single messenger RNA sequence can be translated by multiple ribosomes simultaneously, like an assembly line, each producing polypeptides for new proteins.

- The life of the newly translated protein then proceeds in many different ways, depending on the role that the protein will play; the polypeptide is folded into a unique three-dimensional structure and then transported to the parts of the cell where it is needed. The

general process of translating and transcribing a gene is known in biology as gene expression.

## Suggested Reading

Reece, Urry, Cain, Wasserman, Minorsky, and Jackson, *Campbell Biology*, chap. 13–16, p. 248–324.

Watson, *The Double Helix*.

## Questions to Consider

**1.** What are the implications of labeling a disease "genetic" when there isn't a one-to-one relationship between a gene and the development of the disease? That is, many diseases have a genetic component—a sequence of alleles that makes you more or less likely to develop the disease—but having the genetic predisposition is not a guarantee of getting sick. What are the societal consequences of calling such diseases or conditions "genetic"?

**2.** Some scientists suggest that DNA and RNA chains predated the first living cells—that is, life can be traced back to bits of self-replicating genetic material. If that's the case, does our genetic code belong in our very definition of life?

# Epigenetics, Mutations, and Gene Insertion

## Lecture 6



Each of us has certain genetic predispositions, and genes are the basis for the constellation of traits that make us who we are, but the environment and our experiences impact not only how we use our genes, but even the way that our genes are expressed. As you will learn in this lecture, there are at least three major ways by which our genes change across our life spans: chemical modifications called epigenetics, sequence modifications called mutations, and the insertion of new DNA as a result of viral infections.

**Epigenetics**

- Genes consist of a twisting double helix of DNA sequences. The DNA is not floating freely, but is tightly wound around spools of protein called histones at regular intervals. The combination of histones and DNA creates something like beads on a string, with the beads being nucleosomes. The strings of beads are then packaged and tied by other proteins, and then they are wrapped and packaged again to create the wobbly, slim X shapes of chromosomes.

- The dense packing of DNA within chromosomes saves a tremendous amount of space, but it also serves another purpose: limiting the access of transcription machinery—the tools that turn genes into proteins. If the cellular transcription machinery cannot bind to the DNA, then the genes cannot be expressed—the proteins can't be manufactured. This is not a good or a bad thing, but a way of controlling whether or not a particular gene is expressed at a particular time.

- Gene expression starts with enzymes in the cell called histone acetyltransferases (HATs), which are responsible for linking an acetyl group, a small chemical addition, to the histone protein, which is the spool around which DNA is wound. HATs tack on the

acetyl group through a process called acetylation, which leads to the activation of the genes in close proximity to that histone.

- Histone acetylation relaxes or loosens the DNA structure around the nucleosomes, a combination of histones and DNA, through a process called decondensation. This process facilitates the transcription of genes by enzymes called RNA polymerases, which are responsible for the DNA transcription process. There are natural foils to the HATs called histone deacetylases (HDACs), which remove the acetyl groups and, in doing so, promote the return of the gene to a silent state.

- Although the code from which the protein blueprints are read is relatively stable, if there are no HATs lying around, then the DNA structure can't be loosened enough to read off the instructions. This is one epigenetic way in which the internal environment of the cell can affect the expression of genes.

- A second form of epigenetic change is called DNA methylation, the process of linking molecules called methyl groups to the DNA nucleotides themselves. Note that this is subtly different from histone acetylation, which is the modification of histone protein, not the DNA itself. There are enzymes to help us with this task, and they're called DNA methyltransferases (DNMTs).

- DNMTs seem to suppress gene expression, by physically impeding the binding of enzymes to the gene, for example—or via the actions of proteins that recruit other proteins, including HDACs, which act to silence gene expression.

**Mutations**

- Epigenetic changes are reversible modifications of DNA that do not change the underlying genetic sequence. Changes in the genetic sequence itself do occur, however.

- The cells in our bodies are constantly replicating and renewing our tissues. Each time a cell replicates, it must make a copy of its DNA.

Cells aren't perfect, so occasionally, during the replication process, errors occur, permanently changing the genetic code in the new cells and all future cells born from it. These changes to the genetic sequence are called mutations.

- Mutations can have a range of effects on how the gene is read and how the gene product functions. And, as we age, these changes accumulate, because we've had more opportunities to introduce mutations during cell replication.

- Some mutations occur naturally, as a result of imperfect replication processes during cell division. In other cases, substances found in our environment, such as excessive UV radiation or carcinogens, can damage DNA by breaking it or cross-linking disparate sections.

- Our cells are constantly dividing, and DNA mutations occur all the time. Thankfully, we have a variety of DNA repair proteins in our cells, constantly monitoring and fixing our DNA. Each cell is equipped with strategies for proofreading replicated DNA and repairing errors. If the DNA errors are too severe, the cell can stop dividing, a process called senescence, or commit suicide, known as apoptosis.

- Unfortunately, in some cases, the mutations are undetected or uncorrected, and the genetic alterations have the opposite effect, supercharging aspects of the cell's physiology and promoting uncontrolled proliferation. In essence, this is how cancer develops.

- For example, some mutations might cause uncontrolled gene expression, so the gene product is constantly produced instead of shutting off when it's no longer needed. If the resulting protein contributes to the cell's internal control on cell division, it might promote unregulated replication.

- Other mutations may lead to an overabundance of receptor molecules on the cell's surface. The extra copies of these receptors

endow the cell with extra sensitivity to circulating growth factors or hormones, further promoting their proliferation.

- Mutations can also silence tumor suppressor genes or promote the expression of cancer-inducing genes, called oncogenes.

- Still other mutations can alter the expression of cell proteins that the body normally uses to detect aberrant cells, in effect shielding the cancer cells from detection by our immune system.

- Today, it is firmly established that cancer is a genetic disease, a disease caused by the occurrence and accumulation of an array of genetic alterations that transform normal, healthy cells into malignant cells. Note that calling it a genetic disease doesn't mean that the environment has no influence; the environment of the cell has tremendous power.

- Cancer is still complex and deadly, but advances in genetics are leading a revolution in the understanding and treatment of cancer, so there is much to be hopeful for.

**Gene Insertion**

- Yet another way in which heritability is more subject to change than you might think is through gene insertion. Viruses have been inserting genes into the human genome throughout our evolutionary past, to the extent that almost 10% of our DNA is of viral origin. These bits of viral DNA are called endogenous viral elements, and they are the consequence of infection by a class of viruses called retroviruses.

- Retroviruses are viruses that replicate by turning RNA into DNA, a process that is opposite to the DNA-to-RNA transcription that our cells normally perform. They perform this molecular rewind with an enzyme called DNA reverse transcriptase. Human immunodeficiency virus (HIV) is a type of retrovirus.

- The retrovirus inserts its own DNA into the host cell's genome using reverse transcription—this way, it piggybacks on the host cell's ability to replicate genes, because the virus can't replicate itself.

- Obviously, inserting new DNA into the existing sequence will alter the original sequence. If the viral DNA is inserted into the DNA of germ line cells, those cells that generate the eggs and sperm needed for reproduction, then those inserted mutations will be transferred to the next generation, locking them into one's genetic lineage.

- Most of the viral DNA in our bodies is noncoding at this point, meaning that it is no longer used to produce viral proteins; it's just excess genetic baggage. It used to code for more proteins when our ancestors were first infected, but thankfully, over time, random mutations and our ability to repair DNA have rendered most of it unusable.

- Viruses aren't the only source of gene insertion; human beings have also gotten in on the act. Perhaps the most amazing and controversial aspect of modern genetics is the growing toolbox of techniques used to specifically create or isolate and then insert genes into living organisms. This is taking genetic modification to an entirely new level, a level that allows our species to trump evolution and natural selection and influence our genes.

- There are numerous monogenetic diseases in which the loss or dysfunction of a single gene causes disease, such as sickle-cell anemia and cystic fibrosis. Wouldn't it be great if we could fix nature's mistake and repair or replace the defective gene? Such gene therapy also holds the potential promise (and some would say peril) of genetically modifying crops and livestock, known as genetically modified organisms (GMOs).

- In essence, gene therapy requires two critical ingredients: the DNA of the gene of interest and a means of inserting this DNA into the

host genome so that it is successfully incorporated and reliably expressed in the cells of the host.

- In biological terms, we need a transgene and a vector. A transgene is the gene that will be traveling to another organism. The vector is an intermediary used to pass the genetic material into the host organism.

- In animal studies, transgenes can be introduced directly into the fertilized eggs of a host species. Some of the recipient egg cells will incorporate the new DNA, resulting in offspring that harbor and express the foreign gene product. This approach is impractical and arguably unethical in humans, so a different strategy is needed. We can't just insert new DNA in an embryo without knowing how it will affect the development of a child.

- Viruses are very efficient at transferring genetic material into host cells. Much to our chagrin, the viruses responsible for the common cold and the flu are highly efficient at invading our bodies, entering specific cell types, and hijacking our transcription machinery to replicate their viral DNA or RNA, churning out additional virus copies.



**Viruses transfer their genetic material into host cells in order to make more copies of themselves.**

- Unlike retroviruses, adenoviruses such as cold and flu viruses do not integrate into our chromosomes, so their effect is short term, not potentially lifelong. As such, they make valuable vectors for inserting desirable DNA.

- Ideally, gene therapy proceeds in three stages: A single treatment delivers the repaired gene, the gene is incorporated widely into the

tissue of interest, and the transgene is expressed for long periods—perhaps even a lifetime.

- In 2012, the first gene therapy in the Western world was approved: an adenovirus-based treatment for a rare but life-threatening disorder called lipoprotein lipase deficiency. Individuals afflicted with this disease are unable to produce an enzyme involved in fat metabolism due to a deficit in a gene.

- The new treatment uses an adenovirus engineered to deliver the DNA for the gene directly into the patient's muscle and generate sustained levels of the needed enzyme. The therapeutic effects have been significant and long lasting. Similar gene therapies are sure to be tested for other genetic defects in the coming years.

## Suggested Reading

Reece, Urry, Cain, Wasserman, Minorsky, and Jackson, *Campbell Biology*, chap. 17, 18, and 20, p. 325–380 and 396–425.

Skloot, *The Immortal Life of Henrietta Lacks*.

## Questions to Consider

1. Because our environment and our actions can affect our genetic code, and there are many variables to consider when we attempt to predict whether someone will develop a disease with a genetic component, how much would you want to know? If your genome were sequenced, would you want to know all the likelihoods of all the potential diseases?

2. Should our genetic code be considered personal property? Should we be able to patent our genes? Do they belong to us, or do we give up any rights to ownership when we hand over a biological sample to be analyzed?

# The Illusion of Coherence—How We See

## Lecture 7

We've evolved to trust our eyes implicitly: Seeing is believing, and we seem to experience the world as one continuous image. But different aspects of vision are processed by different parts of the brain. We don't notice that our brains are taking the world apart and putting it back together again, filling in details where we can't see and making inferences about how things should look based on our past experience. In this lecture, you will learn that the way our brain sees is modular, to a large extent, but our conscious experience seems to be coherent, for the most part.

**The Human Eye**

- Our eyes are the most expressive parts of our bodies. In fact, eyeballs are categorized as brain matter, and the eye is the only part of the brain that we can see without surgical intervention.

- From the front, we can see the clear, bulging region of the eye called the cornea. We can also see the pupil—the round, black circle in the center—as well as the surrounding whites of the eye, or the sclera.

- The most prominent feature is the colored, circular patch of tissue surrounding the pupil, the iris. Pigmentation within the iris gives the eyes their color: the browns, blues, greens, and hazels.

- Moving inside the eye, we find a clear, crystalline lens directly behind the pupil. The basic operation of the eye now becomes clear: Light from the environment enters the eye through the pupil and projects through the lens, much like a camera. The pupil acts as a camera aperture, allowing light to pass through the lens and onto the retina in the back of the eye. Within the retina are special cells called photoreceptors, and these cells are the first step in the brain's processing of vision.


**The human eye takes in light for the brain to process and interpret.**

- Like film in a camera, photoreceptors in the retina have the job of capturing the visual scene. But they must also be able to capture a dynamic scene, one that is changing in real time as objects of interest or the viewer move about, and they have the monumental task of converting images into language the brain understands.

- That language is a sequence of electrical impulses. So, we can think of photoreceptors as analog-to-digital converters; the visual scene is out there, in the real world, happening in real time, but our brains create a version of it, and that's our experience of the world. Photoreceptors process light passing through the pupil and lens, sending out a ticker tape of impulses to describe the scene. The brain's job is to synthesize all of this information, creating the cohesive, continuous visual scene that we experience.

- On closer inspection, the camera analogy for the eye starts to break down. In fact, one of the counterintuitive aspects of the eye is that, unlike a camera, the eye works best when it is moving. The photoreceptors of the retina, the light-sensitive cells responsible for

converting light into neural signals, rapidly lose their responsiveness when activated by constant light. Inside the photoreceptors are molecules called photopigments that react to particles of light.

- In technical terms, the loss of light-sensitivity of photoreceptors is called adaptation, because the eye adapts to what is constant in the visual scene. We're only interested in what's changing.

- We imagine ourselves as able to hold a steady gaze or dart our eyes in the quick, flitting eye movements known as saccades. But it turns out that even when our eyes are stationary, they still make very, very small involuntary movements called microsaccades, which allow our eyes, even when fixed on an object, to avoid adaption of the photoreceptors. Without these microsaccades, the visual image would fade out, as the photoreceptors quickly lose their sensitivity.

- Photoreceptors come in two types: rods and cones. As their names suggest, they have different physical appearances, but also different functions. Cones are responsive to the range of wavelengths of light that we experience as different colors. Rods are minor players in color perception, but they are much more sensitive to light, making them useful during the night.

- Cones are responsible for the high visual acuity we experience when we look directly at an object. That is, when we focus our gaze on something, light bouncing off the object passes through to a special region of the retina called the fovea. The fovea is like an island inhabited exclusively by 8 million or so cones.

- Surrounding our island of cones is the ocean of 120 million or so rods, which make up our peripheral vision. Rods are best at sensing ambient light levels, and they modulate the adaptation to light and dark that we experience when we enter or exit a dark movie theater. Rods are also better motion detectors, which is why we notice brief movements out of the corners of our eyes.

- It's at the level of the retina where we find the first example of how the brain deconstructs the visual world into disparate components—because the colors sensed by our cones and the brightness and motion sensed by our rods are processed by different brain regions.

**The Modular Organization of Vision**

- The color-sensing cones come in three different types, and together, they cover the visual color spectrum from violet/red to greens to yellow, using trichromatic vision (like an old color television). These three types of cones correspond to the wavelength of light to which they are most sensitive: short, medium, and long.

- If these cones aren't working correctly or if a person is born without a particular type of cone, then that person will lose sensitivity to the wavelength of light that makes up a certain color.

- Once our photoreceptors convert light signals into electrical impulses, the signals are sent out over cellular cables that come together and exit the back of the eye via the optic nerve, which travels to a relay station in the middle part of the brain called the thalamus. From the thalamus, the visual information is then conveyed to the next higher center, the visual cortex, located in the back of the brain.

- The part of the retina from which the cables exit creates a blind spot; we all walk around with a little part of our visual field that we can't see, because there are no photoreceptors in the part of the retina that receives light from that region. But we don't notice this blind spot, because our brains fill in the missing information.

- It seems to happen pre-attentively—that is, before we've decided where we will focus—and is likely related to the activity of photoreceptors next to the blind spot and the neurons that get that information higher up in the visual processing stream.

- However it works, though, it underscores an important concept about how our brains function: We take shortcuts, heuristics that

we’ve evolved over many thousands of years, so that we can access more information than our limited brains can process quickly.

- Perception of visual information—that is, the psychological experience of seeing—requires a functioning visual cortex. As in many areas of neuroscience, the study of the effects of brain damage gives us insights into how the brain works.

- And this work also demonstrates how specific parts of the brain can be: Damage to one region might affect a particular ability but leave many other skills intact. For example, a stroke in the front and left side of the brain can impair speech and paralyze the right side of the body, but the patient might still be able to read and write and remember and move the left side of his or her body.

- Decades of study have revealed that the visual cortex is a highly organized collection of cells that are tuned to specific features in a visual environment. In addition, research has shown that the visual system parses, deconstructs, and detects the component parts of a visual scene: movement, luminosity, position, orientation, and color.

- We experience the visual world as coherent, even though the way the brain processes vision is modular. This is the binding problem, and it cuts to the core of the intersection of neural physiology, cognition, and the philosophy of consciousness. How do we experience the world as unitary when our brain distributes tasks to different clumps of cells?

**The Binding Problem**

- In the cortex, visual information is passed in parallel in two main directions: up toward the top of the head to the parietal lobe or over and down to the temporal lobe near our ears. These are known as the dorsal and ventral streams.

- These paths are also known as the “where” and “what” pathways. The dorsal stream is called the “where” pathway because it

processes the position of objects in space and coordinates vision with action. The ventral stream is known as the "what" pathway because it recognizes objects, tapping into our memory to identify what we are seeing.

- Brain damage that disrupts the dorsal and ventral streams produces some profound visual deficits, known as agnosias. Patients with visual agnosia are still able to see, but their visual perception is no longer integrated with the rest of the cortex, making it impossible for them to interpret what they see.

- For example, damage to the dorsal stream, or the "where" pathway, causes a condition called hemispatial neglect, in which the patient ignores everything on one side of the visual field. These patients experience the world as if half of it has disappeared; they draw faces with features only on one side, for example.

- Other damage to the dorsal stream can result in the inability to perceive movement. Patients who have motion blindness accurately see stationary objects, but not objects in motion, so cars speeding down the street appear as snapshots. This disturbance can be particularly frightening.

- In contrast, damage to the ventral stream causes impairment to the "what" pathway. These patients are unable to recognize and name everyday objects. We all experience this forgetfulness from time to time, but object agnosia is much more severe.

- An especially dramatic type of agnosia is prosopagnosia, an inability to recognize familiar faces. It's as though every face you see is that of a stranger. These patients use non-facial cues—such as tone of voice, gait, or clothing—to tell their friends from their foes. And their disability underscores how important faces are in our social interactions and daily life.

## Suggested Reading

Livingstone, *Vision and Art*.

Macknik, Martinez-Conde, and Blakeslee, *Sleights of Mind*.

## Questions to Consider

1. Can we ever know whether our experience of color and other aspects of vision is the same as that experienced by other people? How would we test this idea?

2. Memories seem to have the same binding problem as the visual system, because different aspects of a memory are stored in different parts of the brain. How does this segregation affect your experience of remembering?

# Acoustic Perception Deconstructed

## Lecture 8

We live in a rich and complex sensory world. Perception is not just about seeing and hearing what's around us; arguably, a far more important function of our senses is to tune out what's not important. And that's where the modular nature of the brain comes in: We've evolved to pick out what's important for survival and create a personalized experience of the world in our minds. And, as you will learn in this lecture, with no sense do we do that more than with hearing.

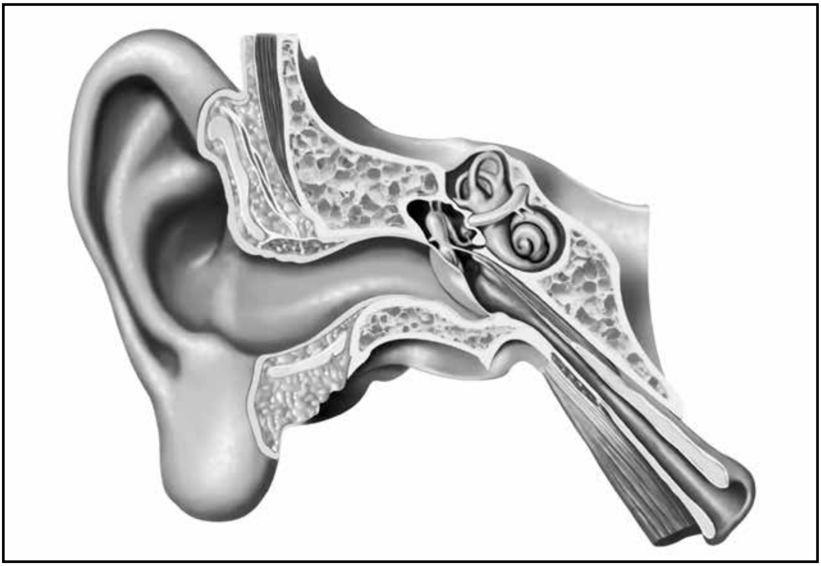
**The Human Ear**

- Our ears are crescent-shaped knobs fixed to the sides of our heads, each fashioning an inner fold leading to a mysterious tunnel. This outer region is called the pinna. Pinnas act as sound collectors, capturing and funneling sound waves into the ear for better hearing. Pinnas are helpful for localizing the source of sounds in space as well as for our sensitivity to sounds that are very faint.

- The little tunnel in our ear is called the ear canal, which leads us to the middle ear, ending at the eardrum. The eardrum is a circular membrane sheet that fills the circular area of the ear canal. Like a drum in a big band, the lightness and tension of the eardrum membrane makes it sensitive to changes in air pressure.

- Attached to the back of the eardrum are three of the most extraordinary bones in our bodies: the malleus, incus, and stapes—which are better known as the hammer, anvil, and stirrup, respectively. They are exquisitely fine and small bones. Anatomists refer to them as ossicles, and because they fit together to connect the middle ear with the inner ear, they are also called the ossicular bridge.

- The hammer, anvil, and stirrup link the eardrum with another membrane called the oval window. Their purpose is to transmit the vibrations occurring at the eardrum to the oval window. As a unit,



© Leonello/iStock/Thinkstock.

**The human ear contains tiny bones that connect the middle ear with the inner ear, transmitting sound along the way.**

the ossicular bridge has joints and flexes, providing a mechanical advantage and higher fidelity than if the sound waves were received directly by the oval window.

- In addition, tiny muscles acting on the ossicles can apply and release tension on the bridge, dampening or heightening the vibrations transmitted from eardrum to oval window. The ossicular bridge is a complicated receiver and amplifier of sound.

- The complexity of these three bones suggests that it likely took a while and many small adaptations for them to evolve into their current state. In order to hear the crunching of leaves beneath the feet of a lion, or the movement of a snake in a tree, our middle ear evolved the ossicular bridge as an amplification system.

- Because the oval window is about 20 times smaller than the eardrum, a tap with the same force makes a much bigger impact on

the oval window than on the eardrum. However, the oval window can also be fairly sensitive to loud noises—so, in response to loud sounds, the muscles that control the tiny ear bones can dampen the force and protect the oval window.

- The oval window marks the boundary of the inner ear and the cochlea, which is our organ for hearing. It looks like a fluid-filled wind sock, curled up like a snail's shell. Just like the ossicles, the shape of the cochlea is the secret to its success.

- The oval window sits at the end of the cochlea, which contains many thousands of fine sensory organs called hair cells. These hair cells are sandwiched between two membranes, or thin coverings, and run the length of the center of the cochlea, as it spirals inward. They are called hair cells because they look like hair; they have a cell body and then a series of little projections (the "hairs") that stand up straight out of the cell.

- In its simplest form, sound is processed in the cochlea when the vibrations from the eardrum are transmitted to the oval window, which then transmits that kinetic energy through the fluid in the cochlea. The fluid motion then causes the membrane that holds the cell bodies of the hair cells, the basilar membrane, to move.

- How the mechanical motion of the other membrane, the tectorial membrane—sitting above the hair cells—affects hearing remains a bit of a mystery.

- When the cell bodies move, the "hairs" bend in relation to the cell body. Motion of the hairs in one direction opens specific gates in the cell, which let in negatively charged molecules, while motion in the other direction makes the cell more positively charged.

- Like the photoreceptors in the eye, the hair cells convert the sensory stimulus, the sound wave, into the language of the brain, electrical impulses, which enter the brain via the auditory nerve.

- An extraordinary feature of the cochlea is the tonotopic layout of the hair cells, in which sound frequency is mapped to the physical arrangement of the cochlea. Real-world experiments have revealed that higher-frequency sounds vibrate the basilar membrane, which holds the hair cells, at the narrowest and stiffest part, the base, whereas low-frequency sounds vibrate it at the apex, where it is the widest and most malleable.

- The reason is that shorter wavelengths peak earlier, at the base of the cochlea, whereas lower frequencies have longer wavelengths, so they cause hair cells closer to the apex, or the wider part of the membrane, to show the biggest responses.

- Then, the auditory nerves that are reading these responses can send messages to other parts of the brain that distinguish between low and high frequencies based on which parts of the cochlea showed the greatest change.

- Most of the time, we find ourselves in an environment with a complex spectrum of sounds with many composite frequencies. Just as the visual system breaks down the visual scene into features, the cochlea parses complex sounds into their constituent frequencies. Then, underscoring the binding problem, the auditory cortex somehow puts the parts back together, and we hear one coherent sound.

**Hearing Loss**

- Perhaps nothing demonstrates the subjective nature of hearing, and the delicate complexity of the hearing process, like the experience of hearing loss, which is the most common physical disability in the United States. Age is the most significant risk factor for hearing loss, with 30% to 40% of people over the age of 65 and half of all people over the age of 75 experiencing some difficulty hearing.

- Our ears are so sensitive that the hair cells in the cochlea can detect a movement of fluid that pushes them the width of a single hydrogen atom. But, over time, we lose this sensitivity, first at the highest frequencies, so speech begins to sound muffled, and high-

pitched sounds like the letter T sound like lower-pitched consonants like the letter F.

- Hearing loss and deafness can occur for several reasons, but at the level of sensation, it can occur with damage to the hair cells in the cochlea. Sustained exposure to loud noises can lead to the death of hair cells, resulting in permanent hearing loss.

- As our culture cranks up the volume in many facets of life, hearing loss is becoming evermore common. We should pay much more attention to the sounds that we allow our ears to endure, because hearing loss resulting from loud noises is almost entirely preventable.

**The Subjective Nature of Hearing**

- Still another way to grasp the subjective, modular nature of hearing is by considering pitch. Pitch is entirely subjective. It has no single physical dimension in the outside world that completely defines it. The same frequency can sound as though it has different pitches, depending on certain characteristics.

- You can discover this conundrum for yourself by taking a tuning fork, or another instrument that can reliably recreate a relatively pure tone, and moving it closer and farther away from your ear.

- Lower-frequency vibrations will sound lower in pitch the closer they come to your ear. Higher-frequency vibrations will have the opposite effect: Closer to your ear, they will sound higher, but as you move it farther away, the sound will get softer, but the pitch will also get lower. This is called the Doppler effect. We often experience it when a car whizzes by us.

- It's not just the Doppler effect, however, that messes up the relationship between sound waves and pitch: Background noise can also affect pitch perception, as pitch seems to be relative. Among a cacophony of low-frequency sounds, a tone will sound higher than if it was embedded in a background of high-frequency noise.

- How is this possible? It turns out that producing a pure tone, one with no frequencies other than the fundamental frequency, is next to impossible. Our everyday experience of sound is as a complex waveform. Even your most accurately tuned piano produces not only a fundamental frequency, with each depression of a key, but also a series of well-defined harmonics, corresponding to the resonant frequencies of the different components of the instrument: the string, the wood, and so on.

- The frequencies that generally accompany a fundamental follow a fairly specific pattern—so much so that if you just played the harmonics of middle C, and eliminated any sound at the actual fundamental frequency, your brain would still interpret the sound as middle C.

- Once again, we have an example of how the brain fills in aspects of the world to help us make sense of it. Just as in the case of the blind spot in our visual field where the optic nerve exits the retina, the brain fills in the fundamental frequency, because most of the time, the fundamental is what lays the groundwork for the complex sound pattern.

- Given that the experience of sound is a creation of each individual's mind, it's not surprising that some people prefer some sounds over others. Each musical instrument produces a unique signature of harmonics, and this signature is what allows us to distinguish a piano from a violin from a voice and also to tell the difference between one violin and another, for example.

- This signature, or characteristic sound, is called the timbre, and each of us will have a subjective preference for different timbres. That's why some people prefer one singing voice and others prefer another.

- Some people are better at distinguishing pitches than others. Most people can recognize a familiar melody, regardless of whether it's being played in the usual key—that is, the pitches are as written by

the composer—or if it's been transposed into a new key or played by a different instrument. The actual notes are different, but the melody is easily distinguished.

## Suggested Reading

Horowitz, *The Universal Sense.*

Test your hearing: http://www.phys.unsw.edu.au/jw/hearing.html or http://www.youtube.com/watch?v=VxcbppCX6Rk.

## Questions to Consider

1. Do you have trouble hearing conversations in certain environments, such as loud restaurants? Have you found any strategies (i.e., sitting with a wall behind you) that seem to help? Why do you think they are helpful?

2. If some hearing loss is inevitable as we age, can our changing musical tastes be a reflection of this loss? Do you prefer different music now than when you were younger? What are the characteristics of the different genres, and how might they relate to your ability to hear different frequencies?

# Our Changing Brain

## Lecture 9

Even up until the last few decades, the brain was thought of as a machine: Once its parts were fully developed in young adulthood, it was fixed for life—of course, with a gradual decline in function as our bodies age. However, in modern times, the idea that the brain is plastic, meaning that structural and functional changes are a fundamental property of the organ that gives rise to our minds, has caused a paradigm shift in the field of neuroscience. As you will learn in this lecture, we call these changes neuroplasticity: We think of the brain as plastic, or malleable, rather than static and rigid.

**Synaptic Plasticity**

- One of the ways in which our brains demonstrate plasticity is with the very smallest changes, which also happen to be the ones that have been studied the longest. These are changes that occur at the synapse, the gap between one brain cell or neuron and another. Brain cells exchange information by sending electrical and/or chemical messages across the synapse.

- A neuron is made up of three basic parts: the dendrites, the soma, and the axon. Dendrites extend out from the rest of the neuron like the branches of a tree, and they receive information from thousands of upstream cells.

- This information from the upstream cells is sent down to the cell body, or soma, where the nucleus, which holds the cell's genetic material, and the other specialized parts of the cell, or organelles, carry out the metabolic and other functions to keep the neuron going.

- The axon is a single projection of the neuron, usually a long, slender nerve fiber whose job is to pass along the information in the form of an electrical signal. Sometimes axons transmit data a very short distance, but they can also reach very long distances.



**Information is sent and received from one neuron to the next so that information can travel between the brain and other parts of the body.**

- The synapse is the junction between cells over which information is exchanged. An axon of one cell usually meets the dendrites of another at the synapse. But a synapse might also link a neuron with either an upstream sensory cell that has information about the outside world or a downstream motor or other cell, whose activity is controlled by the neuron. There are also synapses between dendrites of two neurons or between two cell bodies, but these are not as common.

- At the synapse, the electrical signal (called the action potential) from the upstream (or presynaptic) neuron triggers a cascade of chemical reactions that can have different effects on the downstream (or postsynaptic) cell. The electrical signal causes a voltage change that opens a series of gates, called ion channels. These gates allow the passage of neurotransmitters, signaling molecules located inside the axon to leave the axon and enter the synapse.

- In the synapse, the neurotransmitters float over to the other cell's wall and act like keys in a lock, binding to and opening gates in the

wall. This initiates another series of electrical or chemical changes in that cell that affect the functioning of the downstream neuron.

- Some neurotransmitters excite the cell, triggering an electrical impulse, which we say causes the cell to "fire." And some inhibit the cell, or prevent the cell from firing, or from sending an action potential down its axon. The presynaptic cell turns its postsynaptic partner either on or off.

**Memory Formation**

- Forming memories is how we learn new things, and it is one primary example of how synapses change. The simpler types of memory include learning that two things are associated because they occur at the same time, or one predicts the other. An example is the memory that opening the fridge door will likely lead to the satisfying feeling of eating something delicious.

- Thanks to Russian physiologist Ivan Pavlov and the dogs he trained, we now talk about Pavlovian, or classical, conditioning when referring to the association of a previously neutral object with a stimulus that triggers an automatic reflex, such as a bell that becomes associated with food, which triggers salivation in dogs. This type of reward-based conditioning still remains the most effective way to train your dog.

- In the 1960s and 1970s, neuroscientists Terje Lomo and Tim Bliss discovered that when two neurons that are connected by a synapse are stimulated at the same time, they become associated. This means that in the future, when one fires, the other is more likely to fire as well. We say that the synapse has been "strengthened," because the connection between these two neurons is stronger.

- Let's say that there's a set of cells in your brain that turns on, or fires, when you see the refrigerator door. And there's another group of cells nearby that fires when you see any type of food. Now, whenever you open the fridge door, both sets of cells fire because you see the door and you see the food at the same time.

- Over time, these two cell groups fire together often enough that the connection between them becomes stronger—at the synapse. Just seeing the fridge door is enough to set both your food and fridge detector cells firing. The activity of these cells is now connected, and the fridge door is associated with food.

- This type of association, which can persist for weeks, months, and even years, was first discovered in the rabbit hippocampus. The hippocampus is responsible for turning short-term memories, which last only seconds or minutes at most, into long-term memories that we can recall over the course of a lifetime even.

- The facilitation of the connection between simultaneously active neurons is called long-term potentiation (LTP), because neurons become associated by making it more likely that if the upstream cell sends down an action potential, it will successfully activate the downstream cell, even over the course of weeks, months, and years.

- It turns out that the mechanisms underlying synaptic plasticity are different in different parts of the brain. Even within the hippocampus, the molecular mechanisms underlying LTP can change: The neonatal and mature hippocampi use different proteins and enzymes to accomplish LTP.

**Cortical Mapping**

- We can observe plasticity at the level of the synapse, the tiny gap between nerve cells, but we can also see major changes with experience even in adulthood across large swathes of the brain.

- Often, neuroscientists talk about maps: brain mapping, cortical maps, remapping, and so on. The analogy of a map is particularly appropriate when we talk about the brain because we've learned that there is some topographical organization in terms of brain function: Different functions are localized in different brain regions.

- But the brain is also highly interconnected; no region acts alone. Some regions are more important than others with respect to

certain functions, such as if you have a stroke in the front of your left hemisphere, you'll likely experience some difficulties with language. But those regions are also connected to your memory stores, your decision-making cells, and various other parts of the brain.

- There are two regions of the brain, though, that have been very well charted and seem to be pretty consistent from one person to another, and even between humans and other primates. These are the somatosensory cortex and the motor cortex, those parts of the brain that are responsible for your sense of touch and your ability to control the movements of your body parts.

- These two regions are located near the middle of the top of your head. They take the form of two strips, on either side of a very obvious landmark called the central sulcus. It looks like a large canyon that separates the front from the back of the brain. The motor cortex is closer to the front of your head than the somatosensory cortex, which lies behind the central sulcus.

- What's fascinating about these two strips is that the many parts and functions of the body to which they correspond can be found lying right next to one another, and the more fine motor control or sensation that a body part requires, the more cortical real estate it occupies.

- When you work out or spend years training a particular skill, your homunculus changes as a result. The sensory and motor regions of the brain are plastic and change with use. On the other side, if you don't stimulate a brain region or practice a skill for a long while, the brain will rewire to represent whatever it is that you *are* using it for.

- There's a difference between retaining a basic skill and a high level of proficiency in a complex one. The more complex the skill, the more subtle and specific the wiring in the brain that underlies that ability. Unless you keep up your practice of bike racing, or learning a foreign language, or playing the violin, your ability will decline

as the brain regions that changed as you developed the skill are co-opted for other purposes.

- The plasticity of our brains is also what makes bad habits so hard to break: The more often you do something, the more ingrained the wiring becomes for that particular activity. In fact, it's harder to unlearn bad habits than to learn new ones.

- But by the same token, with appropriate training and effort, we can mold our brains at any age to function more efficiently and more creatively, and we can develop skills that were previously thought impossible. We don't yet understand all the mechanisms of plasticity, and as a result, we don't know what the limits are of our ability to change our brains. Perhaps our capacity to improve is, in fact, limitless.

## Suggested Reading

Doidge, *The Brain That Changes Itself*.

Merzenich, *Soft-Wired*.

## Questions to Consider

1. Why might it be adaptive for the brain to become less changeable and more fixed as we reach adulthood? What would the consequences of an overly malleable brain be in our current cultural climate?

2. What are some examples of conditioning—the equivalent of a dog salivating to a bell with the expectation of receiving food—in your life? Which habits are the hardest for you to break, and how do you think they developed in the first place?

# Plasticity, Brain Training, and Beyond
## Lecture 10

Because the idea that the brain is dynamic and changeable through the life span is relatively new to neuroscience, our understanding of the implications and the potential of neuroplasticity remains quite primitive. And there are certainly no easy ways to improve our minds. Many of the brain-training exercises and tools that are on the market have not been shown to be effective scientifically, but this area of research is one of the most fascinating and groundbreaking subjects in all of science, let alone neuroscience.

**The Plastic Hippocampus**

- Some of the first evidence that the brain can reorganize itself is, in fact, fairly old: Patients who lose brain function as a result of injury or stroke can often regain some of the skills that they lost with time and rehabilitation.

- Ironically, the view that the brain is static in adulthood was largely influenced by the experience of adults with brain trauma: Certain functions can be lost forever as a result of injury to the brain, and recovery, if it happens, is very slow.

- And from what we've learned by studying both plasticity and patients with traumatic brain injuries, the brain is not universally dynamic: Some regions are far more changeable than others, and their malleability is directly related to the functions that they carry out.

- The hippocampus, which enables us to form long-term memories, is one of the most plastic parts of our brains. The hippocampus is also one of the very few regions in the human brain that is not only capable of rapid rewiring, but can actually grow new neurons in adulthood.

- For a long time, neuroscientists thought that all the neurons that form our brains are produced in the womb and that brain development after birth largely involved the forming of connections across the brain, pruning of unattached cells, and various other changes.

- Neurons do not divide the way other cell types do. New neuron generation has now been observed in humans. The growth of these new brain cells is restricted to only two regions of the brain: the olfactory bulb, responsible for your sense of smell, and the dentate gyrus, a subregion of the hippocampus, responsible for turning short-term memories into long-term ones.

- In addition, the neurons that are born are fairly small and make connections with only local neurons; new neurons that project to areas farther away in the brain have not been observed. This is probably a good thing—in fact, very likely a highly adaptive thing—because the migration of neurons is among the most sensitive times in brain development, and we don't want to mess up the wiring in a highly complex system.

- It remains a bit of a mystery, however, as to why, when severed, our major nerve tracts never recover function. Not all vertebrates have this problem: Some species of fish and frogs, with which we share an ancestral past, can regenerate severed spinal cords and other nerves. Even some of our peripheral nerves have the capacity to regenerate and heal.

- So, why are those in our central nervous system, spinal cord, and brain not able to regenerate and heal? The answer might be, in part, related to the location and environment of these nerve cells. It turns out that myelin, the fatty covering that sheathes axons in the central nervous system to speed conduction of electrical signals, also inhibits axonal growth.

- Myelin's job is to make action potential propagation more efficient—to help neurons signal to one another faster, so that

information can travel around the brain and spinal cord as quickly as possible.

- We don't know exactly why myelin also prevents axonal growth, but we can speculate that it might have something to do with the fact that the complexity of the mammalian brain makes it important that, once established, the connections between brain regions and neurons remain fairly stable. Otherwise, we might end up with a mess of wires and a total malfunction.

- The hippocampus, being one of only two areas where new neurons can grow, is also often the culprit when electrical activity in the brain goes awry. The architectural design of the hippocampus, with respect to how its neurons are interconnected, enables rapid and long-lasting synaptic plasticity. This plasticity underlies our ability to form new memories quickly and efficiently, but it also makes the hippocampus prone to starting epileptic seizures.

- The activity that causes a seizure is strikingly similar to the mechanism that allows the encoding of long-term memories—namely, long-term potentiation. Memory, requiring dynamic changes in the hippocampus, and epilepsy, which happens when those changes go unchecked, are closely linked.

- One of the most amazing things that we humans can do is remember something—just about anything—that only happened once. And we seem to have a limitless capacity to remember single events.

- Being a major site of long-term potentiation, and having the ability to grow new neurons, the hippocampus is a factory of neuroplasticity. And besides laying down new long-term memories, the hippocampus is also responsible for our ability to navigate through space.

- It's no coincidence that navigation and memory are related. We even tend to believe our memories for events are more accurate when we can remember where the event happened.

**Training Our Hippocampus**

- The plasticity of the hippocampus is no longer in doubt. Successful training can result in structural changes in the hippocampus. Spatial navigation training can help you grow a bigger hippocampus, for example. What types of training have been shown to be effective by neuroscientists?

- Training has a trajectory. First, more neurons are recruited as we learn a new skill. Over time, though, we become more efficient, using fewer movements, and our neurons reflect that difference by increasing their processing speed and assigning more specific roles to individual cells.

© Ktsimage/iStock/Thinkstock.

**Some parts of the brain, including the hippocampus, change with learning.**

- One of the key features of training, as research has shown, is that you have to pay close attention to the task at hand. And you have to remain motivated.

- Babies, who are little learning machines, have an attention span that seems to be incredibly short. So, why do they learn new things so easily? It turns out that one of the features of the critical period—that time in life when learning seems to be on steroids—is related to attention.

- During the critical period, our brain maps are shaped and refined by experience. And if you're a toddler, everything is new, and you can't know a priori what's important and what isn't. So, your brain is in a state of perpetual knowledge gathering. As your brain develops, you begin to learn what's worth paying attention to, but in the beginning, everything is interesting.

- The key to this heightened learning state seems to be the brain-derived neurotrophic factor (BDNF), which is a protein that is

active in the hippocampus and cortex as well as a region in the front of the brain called the basal forebrain.

- The hippocampus and parts of the cortex show plastic changes with learning, and the nucleus basalis, part of the basal forebrain, is one of the key structures involved in paying attention.

- BDNF is one of the most important proteins regulating the growth of new neurons and synaptic plasticity in the form of long-term potentiation. When neurons fire, BDNF helps them connect. It also promotes the growth of the myelin sheath, the fatty coating that speeds conduction of electrical signals down the axon.

- In infancy, during the critical period, BDNF activates the nucleus basalis and ensures that it stays on the entire time. Some neuroscientists suggest that it puts the brain into its most plastic state and leaves it on until the critical period ends. Then, when the critical period is coming to an end, BDNF shuts down the factory, ensuring stability and decreasing plasticity in the brain.

- How do we put our brains into the plastic state if we're not in the critical period? We need to pay attention effortfully when we want to learn something new and keep BDNF circulating in our system.

- BDNF has also been shown to be released by physical activity, underscoring the importance of remaining physically active in adulthood—not only to keep the body in shape, but also to keep your brain open to learning.

- Often, in our adult lives, we reach a plateau in terms of learning new things. We get out of practice, and our brains become less plastic as a result. Some of the brain training games available on the Internet or via computer are designed to keep us learning. But if you plateau, simply playing a game over and over again won't have much of an effect. The best games are the ones that capture your attention and increase their difficulty as you become more proficient.

- The biggest problem with brain-training software is that it rarely shows transfer: You might get better at the specific game, but transferring that improvement to activities of daily living is rare. If you're interested in such tools, be sure to check how much of the software is based on actual neuroscientific evidence showing it to be effective.

- And don't underestimate the tried-and-true ways of maintaining a high level of cognitive functioning, including getting plenty of exercise and learning a language or new skills that are complex. Fancy computer games are often not as effective as something as ancient as chess.

**Artificial Intelligence**

- If we could implant a device that puts our minds into the critical periods again, would we choose to do so? How would humanity change if our ability to learn was limitless or hundreds of times more effective than it is now? Would such a discovery blow apart the ever-widening gap between the haves and the have-nots?

- We've moved away from thinking about the brain as a machine because it's difficult to imagine machines that are as dynamic and changeable, physically, as our brains. But we've also started to use machines to interface with our brains and enhance their function.

- The development of brain-machine interface devices is a particularly active and exciting area of neuroscience and, in some ways, represents the ultimate plasticity. If we could enhance our brains ourselves, by adding tools to help us think, we might approach a limitless intelligence.

- Currently, in some extreme cases of patients with few other options, tiny computers are implanted into the brain to allow their users to communicate directly with a machine outside of the body simply by thinking about it. This truly might be the advent of extrasensory perception (ESP).

## Suggested Reading

Foer, *Moonwalking with Einstein*.

Kahneman, *Thinking, Fast and Slow*.

## Questions to Consider

**1.** Have you tried any brain-training exercises? Did you find them to be effective in terms of other activities of daily living? Why might the brain act like a muscle, such that training one skill doesn't usually transfer to other skills?

**2.** Why are there only two parts of the brain that grow new neurons? Why are these the olfactory bulb and the hippocampus? Why might our sense of smell and our ability to create new long-term memories require the growth of new neurons, rather than simply rewiring existing ones?

# Magnetism and Its Magic

## Lecture 11

In spite of all we know about magnetism and have been able to do with it, it remains a mysterious force that even the most accomplished physicists don't fully understand. We don't really know why magnets always have a north and a south pole. No one has yet explained why we can't isolate a north or south pole by itself, the way that we can separate negative and positive charges. We also don't really know why particles create magnetic fields in the first place. Understanding these mysteries will occupy scientists for a long time.

**Magnetism: Special Relativity in Action**

- What makes a magnet magnetic? And why are the poles of a magnet called "north" and "south"? The answer lies in the core of the Earth, where the slow movement of iron alloys generates a magnetic field.

- If you let a magnet orient itself, one side will point to the geographic North Pole, and the other will point to the South Pole. The magnetic north and south poles move around a bit because the interior of the Earth changes with time. They're not exactly at the geographic North and South poles—or the top and bottom of the Earth, depending on your perspective. And every few hundred thousand years, the poles reverse.

- But despite these variations in exact position, compasses have been used to help us navigate the world for centuries. And navigation isn't the most important function of the magnetic field: It also protects us from the harmful ultraviolet rays of the Sun. The field deflects most of the charged particles with which the Sun bombards us. If these particles were not deflected, they would erode the Earth's ozone layer, exposing us to more ultraviolet radiation.

- A magnetic field is the area surrounding a magnet's poles that has magnetic properties; a magnetic force is the force that a magnetic

field exerts on a moving charge. We talk about fields in terms of moving charges because magnetism is related closely to electricity, which, simply put, is moving electrical charges—or, even more simply, moving electrons.

© Creatas/Thinkstock.

**All magnets have both a north and a south pole.**

- Magnetic fields are caused by the spin of electrons. Electrons orbit the nucleus of atoms following specific orbital paths. But they also have their own intrinsic spin: They rotate, and this spin is intrinsic in the same way that mass or charge is a property of the electron: If an electron is not spinning, it is no longer an electron. As it spins, each electron generates a magnetic field.

- There are electrons in every atom, so why isn't everything magnetic? In fact, all materials are affected by the magnetic fields generated by their spinning electrons, but most of them contain elements with an equal number of electrons spinning in opposite directions, thereby cancelling out the magnetic fields. Some materials, however, are permanently magnetic; they have a property called ferromagnetism.

- Ferromagnetic materials show magnetic properties even in the absence of a current, or external magnetic field, or remain magnetized even after the external magnetic field is removed. Iron is among the most common examples, but nickel, cobalt, their alloys, and naturally occurring lodestone are quintessential examples of ferromagnetic materials.

- What makes a material permanently magnetic? It's not just the elements that make up these materials themselves; a critical feature of ferromagnetic compounds is their subatomic properties. Without

the bizarre behavior of the subatomic particles involved, magnetism would not exist.

- Ferromagnetism results from two characteristics of quantum mechanics: electron spin and the Pauli exclusion principle. As a result of the field generated by its spin, every electron has what we call a magnetic dipole moment. Basically, it behaves like a tiny magnet.

- Each electron can only be in one of two states: either the magnetic field that it generates points up or it points down. When all or many of the electrons in a ferromagnetic substance are pointing in the same direction, their magnetic fields are additive, and they create a larger magnetic field.

- That only happens, however, in elements whose atoms have a single electron or an uneven number of electrons in an electron shell. Electron shells are essentially the electrons' orbits, or their range of motion around the atom's nucleus.

- If the element's atoms have shells that are filled with pairs of electrons, each electron will be paired with one whose spin goes in the opposite direction from its own, and their dipoles will cancel each other out. As a result, the element won't generate a magnetic field.

- If the element's atoms have single or odd numbers of electrons in their shells, though, the unpaired electrons have nothing to counteract their magnetic properties, and they can be coaxed into alignment with an external magnetic field—or, in the case of certain ferromagnetic materials, they'll align themselves spontaneously. Either way, you get a permanent magnet.

- The Pauli exclusion principle is a principle in quantum mechanics that tells us that two electrons cannot share the same quantum state simultaneously. That is, electrons cannot have the same position in space and the same spin at the same time. They repel each other.

- Inside a ferromagnet, when there's an overlap in the orbital routes of unpaired electrons in the outer shells of adjacent atoms, the electrons are actually farther apart when their spins are parallel, or in the same direction, than when they have opposite spins. And when they are farther apart, the material is more stable, because it has a reduced amount of what we call electrostatic energy.

- Essentially, there is less repellant action between the electrons, or a weaker desire to move away from each other when they are farther apart. So, the ferromagnet holds together after all.

**The Right-Hand Rule**

- The force that a magnetic field exerts on an electric charge is what can cause the charge and the object that carries it to accelerate. Knowing the conditions under which magnetic fields can induce forces is one of the keys to understanding the role of magnetism in electronics.

- To be influenced by a magnetic field, the electric charge must be in motion. The magnetic field cannot get the charge to move if it's not already doing so, and a moving electric charge is what an electric current is. So, in an electronic device, we want to move the electric charge to specific components so that we can power the device or make it function as it's supposed to. In a magnetic field, the charge moves in a direction that is perpendicular to the direction of the magnetic field.

- A magnetic field induces a force, and we need to know the direction of the force to predict how that force will affect the objects that come into contact with the field. The direction of the magnetic field is indicated by a compass needle placed at the point at which you want to find the direction: The field travels from north to south, and the needle of a compass is a permanent magnet whose north pole is at the pointy end and whose south pole is at the opposite end.

- When you place a compass near a magnet, in its magnetic field, the north end is repelled by the north pole of the magnet and, therefore,

points south—which is the direction of the magnetic field. At the northern tip of the magnet, the compass needle will point outward, because the field loops out before turning back to the south.

- Once you know which pole is which, you can use what's known as the right-hand rule to figure out the direction of the magnetic force. To do this, place your hand, palm up, with the fingers parallel to and pointing in the direction of the magnetic field—so toward the south pole, roughly speaking.

- Remember that the electric charge must be moving perpendicular to the direction of the magnetic field, so if you extend your thumb, it will point in the direction of the electric charge. Then, the force applied to a charged particle will be in the direction that would shoot out of the center of your palm.

- This rule is used to find the direction of the magnetic force on a positive charge. If the charge is negative, the force goes in the opposite direction—so toward your palm rather than away from it.

- We can consider an electric current simply as a stream of charged particles and still use the right-hand rule to figure out in which direction the magnetic field would act. But a current also produces its own magnetic field.

- This field follows a circular pattern around the wire in which the current is traveling, and you can use a version of the right-hand rule to find the direction of the field. Knowing the direction of the field—and, therefore, the magnetic force—tells you how the force is being used to move mechanical bits in various electronics, such as the voice coil and the cone that it's attached to in a loudspeaker.

- To find the direction of the field, curl the fingers of your right hand to form a semicircle, and point your thumb in the direction of the current. Now the tips of your fingers point in the direction of the magnetic field.

**MRI Technology**

- One application of magnetism that has changed the field of medicine is the magnetic resonance imaging technology (MRI) that allows physicians to see tissues previously only visible via invasive surgical interventions.

- MRI machines are essentially very strong magnets that take advantage of the fact that our tissues are made mostly of water, with two protons, or hydrogen atoms, for each oxygen atom. Inside the powerful magnet, most of the dipoles of the protons line up with the direction of the field, following the right-hand rule.

- A brief radio-frequency signal is passed through the machine. This frequency is set so that it flips the spin of the protons. Then, the protons are allowed to relax as the electromagnetic field is briefly turned off, and the receiver coils read the radio frequency signal that is generated.

- Protons in different tissues relax at different rates, resulting in different signals sent to the coils. So, for example, radiologists can then tell the difference between brain tissue and bone tissue, or tumor tissue and healthy tissue. Additional magnetic fields can be generated during the scans to create 3-D images, or scans can be taken over time, and then compared, to watch how blood flows through the brain and other tissues.

- MRI scans are very safe, especially when compared to the harmful radiation that results from X-rays or PET scans. The biggest worry is that some metal piece, like a pacemaker or screw, might shift, causing irreparable damage to the part of the body in which it is embedded. But exposure to the magnet and the radio frequencies themselves has not been shown to cause any harm to patients undergoing the procedure.

## Suggested Reading

Fleisch, *A Student's Guide to Maxwell's Equations*.

Livingston, *Driving Force*.

———, *Rising Force*.

## Questions to Consider

1. Why do magnets seem magical?

2. When the poles flip polarity, will we have to adjust our compasses and GPS systems? Why or why not?

# Electrical Forces, Fields, and Circuits
## Lecture 12

Electricity seems to have quasi-magical powers, just like magnetism, its closely related cousin. And yet the world as we know it would not exist if electricity failed to flow through our bodies, our homes, and our cities. In this lecture, you will learn about many aspects of electricity—from electric charges to action potentials and the wiring in your home—and you will come to understand how important electricity is as a form of energy, in our bodies and in the world at large.

**Electricity Basics**

- Every atom with an electron has the potential to create an electric charge; electrons have an inherent negative charge. There are two other particles in the nucleus of atoms: protons, which have an inherent positive charge, and neutrons, which are neutral. Most atoms found in nature are electrically neutral because the number of electrons equals the number of protons in the atom and the charges cancel each other out, leaving the atom with no net charge.

- But electrons can be transferred from one atom to another, disrupting this fine balance. This flow of electrons is what we call electricity. When two different surfaces are rubbed together, electrons can be transferred from one to the other. Electrons get transferred, but no new electrons or other particles are created—so, whereas one substance becomes more negative, the other becomes more positive. The result is static electricity.

- It's not just by physical contact that electrons get pushed around; chemical reactions, electrical circuits, and radioactive decay are just three more ways to change the electrical charge of different objects. But within a closed system, where no particles flow into or out of it, the total electric charge remains static: The law of conservation of electric charge holds that no new charges are created.

- Because like charges repel each other and opposite charges attract, electric forces can alter the motion of an object. For example, if you move two like-charged objects toward each other, the force between them will cause one to repel the other—moving them away from each other.

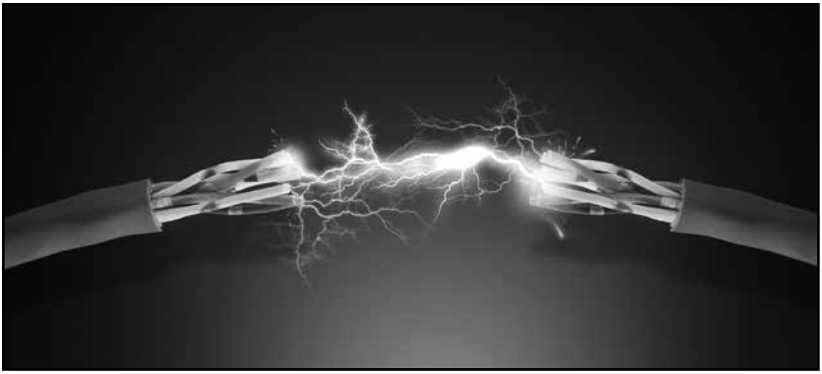
- While electric charge can be transferred from one object to another, it can also move along or through an object. But different materials are more or less helpful in moving along the charge: Some substances are good electrical conductors, like metals, and others are poor conductors, which we call electrical insulators, like rubber, some plastics, and wood.

- What makes one material a conductor and another an insulator is all about the atomic structure. Because the outermost electrons in an atom are the farthest away from the nucleus, they are also the easiest to pick off. In materials that are good conductors, these valence electrons, as they are called, often pop off and hang out freely throughout the substance.

- Good conductors are materials that have more electrons that are easier to pick off. When a negatively charged object comes into contact with the substance, the unruly sailor electrons get out of its way, traveling farther from the point of contact. If a positively charged object is around, they will find their way over to where that object is located.

- In insulators, electrons are kept in line, and there are far fewer wanderers moving about the decks. So, when charged objects come into contact with an insulator, the electrons stay put, and the flow of charge is stopped.

- An electric field is made by static electric charges. When these charges are in motion, the field is electromagnetic, as the motion induces a magnetic field as well. When we illustrate a field, the field lines move away from positive charges and toward negative charges.





**Electricity is the flow that occurs when electrons are transferred from one atom to another.**

## Electric Potential

- A concept that incorporates the idea that more charges mean more energy is the electric potential. The electric potential energy of an object is a combination of its own electric charge and its relative position to other charged objects. The electric potential of an object or system is its electric potential energy per unit charge measured in volts.

- Because like charges repel and opposite charges attract, bringing two like charges together takes work. The same is true of separating opposite charges. So, when two like charges are brought closer together, there is an increase in electric potential energy in that system. And the more charges are involved, the greater the potential energy. Energy is measured in joules, while charges are measured in coulombs.

- The potential difference between two points is measured in volts and is called voltage. Positive charges flow from a region of higher electric potential to one of lower electric potential—that is, positive charges move in the direction of lower electric potential. The opposite is true for negative charges: They go from regions of lower potential toward regions of higher potential.

- All of the nerve cells in our bodies—whether they are located in the brain, spinal cord, or any part of the peripheral nervous system—send and receive information in the form of electrical signals.

- The cells in our nervous system are called neurons, and they come in many different shapes and sizes, but they all have a few characteristics in common: They receive signals through branch-like structures called dendrites; send those into the cell body or soma; and then relay the signal down a single axon, a long, thin protrusion that in turn sends it along to the dendrites of other neurons, or to the receptor region of a muscle cell or other type of cell that uses the information it receives.

- These electrical signals depend on the potential difference at the boundary of the cell, the cell membrane. Ions—that is, atoms or molecules with an uneven number of protons and electrons—abound in the extracellular and intracellular fluid. Because they have uneven numbers of protons and electrons, they hold a positive or negative charge, depending on whether they have more protons or more electrons.

- Outside the cell, sodium ions, holding a positive charge, and chloride ions, with a negative charge, are plentiful, while inside the cell, positive potassium ions and negatively charged proteins fill the intracellular fluid.

- By the process of diffusion, molecules like to move from an area of high concentration to an area of low concentration. So, the tendency is for sodium and chloride to want to move into the cell and for potassium to move out. However, the cell wall is only selectively permeable—that is, only some molecules or compounds can pass through. The membrane achieves this selective permeability by containing channels that let in some molecules but not others.

- Neurons have channels that let out potassium ions easily but make it harder for sodium to pass through. Because potassium leaves and the positive charges aren't replaced immediately by sodium, the

interior of the cell becomes slightly negatively charged near the channels. This negative charge attracts positive ions on the other side of the membrane, so a slightly positive charge builds up on the outside of the cell.

- This separation of positive and negative charges creates an electric potential difference across the membrane, which we call the resting membrane potential. To maintain this resting potential, and the sodium and potassium ion gradients, cells are equipped with sodium-potassium pumps, which require energy in the form of a molecule called ATP.

- For every three sodium ions that the pump dumps out of the cell, it pulls two potassium ions in. These pumps keep more potassium inside and more sodium outside, which results in a slightly negative charge. But it's the ion channels that are specific to sodium and potassium that are responsible for most of the negative charge.

- When a neuron is at rest, or not firing a signal, this membrane potential is negative on the inside with respect to the outside. Then, with the right trigger, a series of events leads to the propagation of an electrical signal, which we call an action potential.

- Some stimulus—whether it's an action potential from an upstream neuron, or a photon in the retina, or the bending of a hair cell in the inner ear—causes some sodium channels in the neuron to open along the membrane and allow sodium ions to flow in freely. These ions are driven into the cell because there are fewer sodium ions there and there is a slight negative charge.

- We call this the depolarization of the cell, because it's going from a polarized, or negatively charged, state toward a more neutral state. If the depolarization is strong enough and reaches a specific threshold, then an action potential is triggered.

- An action potential consists of an initial rise in the membrane potential: It goes from negative to neutral and overshoots the

neutral point to become slightly positive. Then, it quickly returns to its negative state: We call this the falling phase, and it even undershoots the resting potential a little bit—that is, it becomes slightly more negative than the typical membrane potential. Eventually, it returns to its resting state.

**Circuits**

- An electric circuit is a closed-loop system in which electrons can flow to generate electricity. In every circuit, there is some power source, such as a battery or a generator, that produces the pressure or force necessary to move electrons.

- We measure the force in terms of volts, or electric potential energy per unit charge. The electric current is measured in amperes, the amount of electric charge passing a point per unit time. To calculate the total power in the circuit, we multiply the current by the force and get the wattage.

- If you think about an electric circuit as analogous to your own cardiovascular system, then the heart functions as the battery, and the rate of blood flow is the electric current. The force with which the blood flows is the voltage.

- But not all of your arteries and veins are clear of obstacles; sometimes, the buildup of plaque or other compounds impedes your blood flow. The physical structure of the artery or wire can increase or decrease the resistance to the flow. In an electrical circuit, we measure the resistance in units of ohms, which is the voltage divided by the current.

- The larger the diameter of the artery or wire through which the blood or current flows, the less resistance there is. In the brain, the wider the diameter of the axons in our neurons, the faster the action potential travels down to the synapse.

- A circuit can be designed such that the current flows from one area to the next in a serial fashion, like in a string of Christmas lights. If

one bulb burns out, all the bulbs remain dark because the current can't flow. We call this a series circuit.

- But another way to design a circuit is in parallel, like the wiring in a house. If one light bulb goes out, you can still use electricity in other parts of the house, even though they are all powered by the same source. This is because each appliance or bulb or section of the house can almost be thought of as a separate circuit, all of them working in parallel. Current flows into and out of each device or section of the house separately.

- We can be thankful that our brains didn't develop as a series circuit but, instead, as a very complicated set of parallel circuits. Even if we happen to blow one mental bulb, the rest of the lights will stay on.

## Suggested Reading

Kamkwanba and Mealer, *The Boy Who Harnessed the Wind*.

Patrick, *Nikola Tesla*.

## Questions to Consider

**1.** What would our world be like had we never learned to harness electricity? What if everything were powered by electricity, rather than any other energy source?

**2.** Why would the body harness electricity to facilitate communication between cells, especially in the nervous system? What would have been another viable solution for Mother Nature to consider?

# Thermodynamics—Heat, Energy, and Work

## Lecture 13

Thermodynamics, from the Greek words for "heat" and "movement," is the study of heat's relationship to work, and to energy in other forms. This lecture explores the concept of power in machines as predicted by the laws of thermodynamics. In this lecture, you will start by gaining an understanding of the relationship between temperature, heat, and work. Then, you will move on to the laws of thermodynamics and to exploring how engines apply these laws to power our vehicles.

**The Zeroth and Third Laws of Thermodynamics**

- The history of modern thermodynamics is often said to start with a publication by French military engineer and physicist Nicolas-léonard-sadi Carnot, the father of thermodynamics. With his book, published in 1824, titled *Reflections on the Motive Power of Fire*, he ushered in a new scientific discipline, outlining how an idealized engine might function.

- The title of Carnot's seminal work underscores the central relationship that defines thermodynamics—that is, the relationship between heat, or fire, and motion, or how heat powers things.

- Heat is energy that flows from a higher temperature to a lower temperature, often across objects in contact with one another. We measure heat in joules. We don't say that an object contains heat—it has a particular energy, but unless the heat is moving from one place to another, it's not heat. Heat moves from coffee in a mug to the mug itself, and then to our hands as we hold it.

- When the temperature of an object stays steady, and there is no net flow of heat from one area to another, we say that the object has reached thermal equilibrium. We measure the temperature of an object through the application of the zeroth law of thermodynamics, which states that if two bodies are independently in thermal

equilibrium with a third body, they are in equilibrium with each other. In other words, we can use a thermometer to measure temperature and, therefore, predict the flow of heat.

- The third law of thermodynamics states that it's impossible to reach a temperature of absolute zero, which is the coldest temperature. At absolute zero, or at about 273.15 degrees Celsius, all molecular motion stops. The third law states that it's impossible for something to reach absolute zero because heat flows from hot to cold, so the surroundings of something going to zero need to be less than absolute zero, which is impossible.

**The First and Second Laws of Thermodynamics**

- The first law of thermodynamics is that of conservation of energy, though it applies to mass as well: Energy cannot be created or destroyed in an isolated system. This means that energy in a system can only change form—for example, from heat to work, or vice versa—but you cannot add it or subtract it. It has to come from somewhere in the system.

- The second law states that heat flows from hot to cold and never the other way around. To get something to go from cold to hot, then, requires work. The second law is often also stated in terms of entropy, or the state of order in a system. Using this framework, the second law states that a system is always moving toward maximum entropy. Life devolves into chaos.

- With thermodynamics, we're interested in how a particular system works—whether it's an ocean, a car, or a thermometer—so we divide the world into our system and its surroundings. In each case, we have a number of different variables to consider, so when we are analyzing the thermodynamics of an object, we need to be clear on what we're calling the system and what is defined as its surroundings.

- We can also have open or closed systems. In an open system, there is heat or work flowing into and out of the system, such as a dam that is constantly replenished by a water source elsewhere.

In a closed system, there is no movement of heat, work, or energy into or out of it—it's completely isolated. Most systems aren't, but it is sometimes most effective to think of a particular system as theoretically closed.

- In a closed system, we can create what's called an ideal engine. Carnot proposed that a maximally efficient heat engine has one important property: All the processes within the engine must be reversible. That is, when both the system and its surroundings are returned to their original states after the process has occurred, we say that it's reversible.

- We know, then, that a car's engine is not ideal because heat escapes and because processes within the car cause friction, which dissipates energy. Note that the opposite isn't true; a reversible engine may not be maximally efficient.

- Perhaps the best way to imagine a Carnot engine is to think of a system in which heat comes from a hot reservoir, powers the engine, and then rejected heat or unused heat flows into a cold reservoir.

- The engine itself has some unique features: There is no friction, so no energy dissipates that way, and the hot side of the engine is the same temperature as the hot reservoir and the cold side of the engine is the same temperature as the cold reservoir.

**Types of Engines**

- In order to understand how the laws of thermodynamics enable us to power our lives, it's useful to concentrate on two types of engines that are perhaps the most famous examples of thermodynamics in action: the steam engine, in old trains and ships, which is an external combustion engine; and our car engines, or an internal combustion engine.

- Steam engines were the first to be widely used, and improvements to the steam engine ushered in the industrial revolution. And they

didn't just power boats and trains; they were also responsible for powering many factories.

- What are the major components of a steam engine? First, you need a piston attached to a rod and placed inside a cylinder. The movement of this piston is what moves the train, boat, or whatever forward. What causes the piston to move is the hot air in the form of steam. So, you need a place for the excess steam to exit and a place for the high-pressure steam to come in.

- Then, you need a way to heat up the steam—a boiler, which is simply a furnace that heats up the water. There are two main types: In a fire-tube boiler, pipes run hot gases from the fire through a vat of water, whereas in a watertube boiler, which are more common these days, the hot gases heat a series of tubes of water.

- The steam leaving the boiler is at a very high pressure, and when the engineer opens a valve to let that steam travel down to the piston chamber, it causes the piston to move in one direction.

- There is another valve attached to the piston chamber that moves when the piston is all the way at the other end of the chamber. When that valve moves in the opposite direction, it lets out excess steam, which causes the piston to move back to its original position. This way, the pressure from the steam moves the piston back, and the valve moves it forward. This is why excess steam is let out of the chimney.

- Steam engines are external combustion engines because the combustion that generates the heat that creates the work occurs outside of the engine itself. Internal combustion engines are much more efficient and, therefore, can be a lot smaller, making them ideal for vehicles like cars.

- The inventors of the internal combustion engine noticed that much of the potential energy in the furnace that heats the water in a steam engine is wasted. Instead, if you ignite even a small amount of fuel

in an enclosed space, a very large amount of energy is released, and you can use that energy to put something else in motion. Heat is used to create work. That's the fundamental principle underlying a car engine.

- In 1867, Nikolaus Otto invented the four-stroke combustion cycle to convert gasoline into motion. The strokes refer to the movement of the piston up and down the cylinder, which goes through four strokes in each cycle.

- The first one is the intake stroke: The piston moves downward to let the engine take in a cylinder volume of air and a tiny amount of gasoline. Then, during the second stroke, the piston moves back up to compress the fuel/air mixture, ensuring that that explosion will be more powerful.

- At the top of the third stroke, the combustion stroke, a spark plug emits a spark that ignites the fuel and drives the piston down. Then, the fourth stroke opens the exhaust valve so that the exhaust can leave the cylinder and go out the tailpipe.

- A combustion engine uses the principles of thermodynamics to move a car forward by converting gasoline into motion. But there's another type of engine that is far more powerful: the turbofan engine—a type of gas turbine—that powers commercial jets.

- Besides gas turbines, there are three other main types of turbine engines: steam turbines, water turbines, and wind turbines. Coal, natural gas, and nuclear power plants all create steam and run it through a massive turbine to spin some kind of shaft that drives an electricity generator. A turbine is a mechanical device that rotates and extracts energy from the steam to move the shaft.

- Hydroelectric plants use water to power the turbine, while wind turbines are moved by the wind. In a gas turbine, essentially the same concept applies: A pressurized gas spins the turbine, which creates work and drives some component. In a jet engine, heat from

© Carlos Santa Maria/iStock/Thinkstock.

**Jets operate on the basis of Newton's third law of motion.**

burning jet fuel expands air, and this fast-moving hot air spins the turbine, which we see under the wings in commercial airliners.

- Jet fuel is expensive, but the amount of power that you get in a jet engine for its weight, compared to a traditional engine, is much greater. Gas turbine engines are also much smaller than other engines.

- A gas turbine starts with a rotating device with spokes that takes in outside air and compresses it, sending it to the combustion chamber behind it. There, the fuel is burned, and then the turbine, located behind the combustion chamber, extracts the pressurized, fast-moving gases and moves the output shaft.

- In a turbofan engine, the shaft is connected back to the front of the engine, where it powers a big fan—the blades that you see when you look at the front of a jet. This fan dramatically increases the amount of air moving through the engine and increases its thrust, which is what moves the airplane forward.

## Suggested Reading

Carnot, *Reflections on the Motive Power of Fire and Other Papers on the Second Law of Thermodynamics*.

Cengel and Boles, *Thermodynamics*.

## Questions to Consider

1. Someday, electric cars might completely replace combustion engines. What makes electricity a more or less desirable alternative to the combustion engine, in terms of the efficiency of turning energy into work?

2. If a hurricane or a cyclone is a particularly efficient Carnot engine, how does the second law of thermodynamics, suggesting that order descends into chaos, predict its path of destruction?

# Metabolism—Energy in the Cell

## Lecture 14

This lecture turns to the chemistry of thermodynamics to investigate how heat, or another source of energy, can turn into work. How a cell transforms matter and energy to do work is called metabolism. In this lecture, you are going to take a closer look at the power generators in our cells and those of plants. The goal of this lecture is for you to understand how cells turn sunlight or food into a form of energy that can be used to do the necessary work.

**Photosynthesis: Our Solar-Powered World**

- Photosynthesis, or the conversion of sunlight into sugar and other chemical energy stores, provides sustenance for almost every living thing on Earth, either directly or indirectly. Even the fuels that we burn to heat our homes and power our cars have their origins in plant and animal life that put the Sun's energy into storage hundreds of millions of years ago.

- Human cells have evolved without the capacity for photosynthesis. What makes plant cells special? The answer lies in part in their structure. Our ability to move around would seriously impede the ability to capture energy from the Sun the way that plants do. Our cells are squishy and mobile; plant cells are much more stable and structured. This architecture allows them to take advantage of certain properties of a specific molecule: chlorophyll.

- To turn sunlight into energy, plant cells take carbon dioxide molecules from the air and water molecules from the soil and add the energy of sunlight to make glucose, or sugar. Then, they either use the molecules of extra oxygen and water in other reactions or expel them for our consumption.


**In the process of photosynthesis, plants use energy from the Sun to produce food and, as a by-product, oxygen.**

- This reaction is essentially what our cells do—but in reverse. We add oxygen and water to glucose and create the energy that our cells need to work plus the by-products: carbon dioxide and water.

- Photosynthesis largely consists of two components. First, the light reactions capture energy and require sunlight. Second, that energy is then used to build glucose and other molecules in a process known as the Calvin cycle. These are also called the dark reactions, because they don't require light.

- Both of these processes happen in an organelle called a chloroplast, which is unique to plant cells. Chloroplasts are similar to the mitochondria that power animal cells. All eukaryotes have mitochondria, but plant cells also have chloroplasts.

- The structure of chloroplasts is the key to their success. Surrounding the organelle is an envelope of two membranes, like the nucleus in a eukaryotic cell, and inside it is a fluid called the stroma. Within

the stroma is a third membrane system, which is made up of sacs called thylakoids.

- Thylakoids are often stacked in columns called grana. This structure, while highly efficient, simply wouldn't be possible if plants were mobile the way animals are; it's flexible enough to bend a bit, but it is just too rigid for walking, running, and jumping.

- Chlorophyll, which is pigmented green and is the star of the photosynthesis show, is found in the thylakoid membranes. The light reactions convert solar energy to chemical energy, and this happens in the thylakoids, where the chlorophyll is.

- Photosynthesis is the process by which light is turned into food: First, the light reactions capture energy from the Sun in the form of photons and then use it to make adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide phosphate (NADPH). Then, the Calvin cycle, or the dark reactions, uses these products to create sugar with the help of carbon dioxide.

- The sugars then supply the plant with chemical energy and carbon skeletons so that the plant cells can manufacture all the molecules they need to perform the functions of life. The mitochondria in plant cells use the products of photosynthesis as fuel, and the excess sugars are stored in various parts of the plant, like in seeds, roots, and fruits.

- Humans and other animals are attracted to plants: They eat their seeds and fruits and disseminate the plants' genetic material so that new plants can grow, because, of course, plants can't move around by themselves.

- Perhaps most importantly, one of the by-products of photosynthesis is oxygen, which animals like humans need to survive. Plants use up carbon dioxide, which is emitted by animals during their own metabolic processes. Without photosynthesis, life could not survive

on Earth. And metabolism in animals helps power photosynthesis by providing carbon dioxide.

**Cellular Respiration**

- Cells have multiple jobs, and of course, different types of cells have different subspecialties, from muscle control to information signaling to detoxification and development. All of these jobs, and the factories that create the products needed to accomplish them, require fuel. In animals, food needs to be ingested and converted to energy so that the protein factories can be powered.

- This process of converting food into fuel is called cellular respiration, which is photosynthesis in reverse—taking sugars, adding oxygen to them, breaking them down to release energy, and expelling carbon dioxide as a by-product.

- Just like in the internal combustion engine in your car, your cells mix hydrocarbons with oxygen and let out an exhaust of carbon dioxide and water. Moving electrons around releases energy in organic molecules, energy that is ultimately used to build ATP, so that the energy can be stored for later use.

- Cellular respiration can be divided into three main phases: glycolysis, the citric acid cycle, and electron transport. Glycolysis is the process by which sugar, in the form of glucose, is split into two molecules of a compound called pyruvate. It's a 10-step process, and the end products are pyruvate and two molecules of energy-storing ATP.

- After glycolysis, pyruvate migrates to the mitochondria and enters the organelle. There, oxygen is added to a reaction that produces a compound called acetyl coenzyme A (acetyl CoA) that then enters the citric acid cycle.

- Like the Calvin cycle in photosynthesis, the citric acid cycle, or Krebs cycle, can be thought of as an assembly line that manufactures specific products. Acetyl CoA enters, and then there are 8 steps that

eventually result in the release of ATP and a return to the original state so that another acetyl CoA molecule can pass through.

- After the Krebs cycle, the last step in cellular respiration is what is called the electron transport chain, in which electrons are passed along in a series of reactions, gradually releasing energy that can be used to make more ATP.

- The building blocks of the chain are located in an internal membrane within the mitochondria so that the different molecules in the chain can be held closely together. In this way, the structure of the mitochondria is reminiscent of the chloroplasts.

- Ultimately, the electrons are passed down to oxygen, the final electron receptor. Because of its polar properties, oxygen draws the electrons toward it along the chain, and because we breathe in and out all the time, there's plenty of oxygen to go around among all of our cells.

- But the electron transport chain is just the first step in this last phase of metabolism in the mitochondria. In fact, the next step is likely the most important. Whereas glycolysis and the citric acid cycle produce about 2 net molecules of ATP for use elsewhere during each cycle, oxidative phosphorylation, or the combination of the electron transport chain and a process called chemiosmosis, produces 26 to 28 molecules of ATP.

- In the chemiosmotic process, energy from the electrons in the transport chain is used to pump protons, or hydrogen ions, across the membrane, and then they move back through an enzyme called ATP synthase. The flow of protons back across the membrane provides the energy needed to form ATP. The remaining electrons and protons at the end of the pumps are joined with oxygen to form water.

**Metabolic Engineering**

- Knowledge of the metabolic pathways, or chemical reactions in cells, has been used extensively to improve our lives. Harnessing a

cell's metabolism to build products is called metabolic engineering, and the ultimate goal is to develop cost-effective and ethical ways to produce valuable substances in large amounts.

- Current metabolic engineering techniques involve using well-known, safe microorganisms on an industrial scale to produce bulk chemicals such as those used in fragrances, essential oils, vitamins, agriculture, and many more, but the future entails tailor-made designer cells that might make synthetic organic chemistry production obsolete.

- One of the first major success stories involving metabolic engineering is the production of an anticancer therapeutic agent, currently used in the treatment of several types of cancer. In this story, the needs of cancer patients were pitted against the environmental costs of harvesting a naturally made compound.

- A chemical called Taxol was discovered in the 1960s; it was collected from the bark of a Pacific yew tree known as *taxus brevifolia* in a forest in northern Washington state. In the late 1980s, the compound was shown to have a beneficial effect on patients with melanoma, or skin cancer, and ovarian cancer, but in order to treat all the current cases in the United States, the pharmaceutical company would have to destroy 360,000 trees every year.

- By the 1990s, as the environmental movement was rapidly growing in the United States, there was pressure to create the drug synthetically, saving the trees in the process. The best method discovered took advantage of metabolic engineering and, in the process, eliminated the environmental costs of killing trees. It also uses fewer toxic chemicals in its production than purely synthetic methods, and it uses less energy.

- Today, the drug Taxol is FDA approved for the treatment of breast, ovarian, and lung cancer, and it is even available in a generic form. All production of Taxol by the Bristol-Myers Squibb pharmaceutical

company uses plant cell fermentation technology, and there is no longer any need to take bark from trees.

- Metabolic engineering is now widely used in the pharmaceutical industry, but it has also improved production of some of our favorite vices: cheese, beer, and wine.

## Suggested Reading

Field, *Culinary Reactions*.

Pollan, *The Botany of Desire*.

## Questions to Consider

**1.** If we wanted to make our bodies more efficient, can we do so by changing our diet?

**2.** Why do we feel sluggish after a meal, given that our bodies have just been injected with fuel?

# Fluid Mechanics—Pressure, Buoyancy, Flow

## Lecture 15

Fluid mechanics describes the relationship between pressure, mass, and the flow of fluids. The principles of fluid mechanics govern how our blood flows throughout our bodies and how our internal plumbing system functions and fails. In this lecture, you are going to explore these principles and how they govern how a plane is lifted into the air, why a ship floats in the ocean, and how the movement of fluid molecules in air and water allows these things to happen.

**Pressure and Depth**

- When we talk about fluids, we are referring to any material that can flow, and that includes both liquids and gases. These materials differ in their respective states, which we intuitively recognize as the difference in their mass density—the mass per unit of volume, or how many molecules are stuffed into a specified area of space.

- Air has a low mass density because its molecules are spread widely apart, while water has a much higher mass density because its molecules are tightly packed. And solids have the greatest mass density.

- You can think of pressure as the force acting directly on a surface—that is, force exerted at an angle perpendicular to the plane of the surface. Technically, pressure is the magnitude of the force per unit area. Pressure is around us all the time, as in the air pressure of the atmosphere.

- In order to understand how planes fly, we need to become familiar with the forces and the relationships between them that underlie the physics of flight—specifically, pressure in a static environment, or the relationship between pressure and depth.

- In general, as vertical depth increases—in other words, as you go downward in a fluid—fluid pressure also increases, because the lower layers are compressed by the weight of the upper layers. Air is at a higher pressure at sea level than at the top of Mount Everest, the difference in vertical depth being the height of the mountain.

- You might have felt the pain in your ears as you swim to the bottom of the deep end of a swimming pool. The fluid pressure at the bottom of the pool is higher than that at the surface because there is more water weighing it down, and we can feel that pressure increase as the painful push on the outside of our eardrums as we descend.

- The change in pressure swimming to the bottom of the pool is much greater, even with just a few feet of descent, than is the change in pressure when hiking to the top of a mountain from sea level, which could be hundreds or even thousands of feet. The difference is the density of the fluid surrounding us. Static fluid with a higher mass density will accrue a larger pressure difference as the vertical depth increases.

**Buoyant Force**

- Why do some objects float and others sink in a swimming pool? Density is part of the answer. Water has a mass density, and ice has a mass density that is lower than that of water, so ice floats.

- Buoyant force is the force that all fluids apply to objects immersed in them. Objects in a fluid displace that fluid, so you can think of the buoyant force as an opposing, upward force exerted by the fluid back onto the object.

- The Greek scientist Archimedes determined that the magnitude of the buoyant force equals the weight of the fluid that the object displaces. This is known as Archimedes' principle.

- For objects that are solid throughout, whether or not that object will float depends only on its density relative to the density of water and

© Fuse/Thinkstock.

**The concept of buoyancy explains why heavy cruise ships are able to float on water.**

the maximum buoyant force that water can exert. Being solid is the important factor here.

- The steel that makes up a cruise ship will not itself float in water. But steel as a part of a mixed material object, like a cruise ship, can be made to float. The addition of other materials, like all the air within a ship's hull, means that the buoyant force created by the huge volume of water displaced by the massive ship can, and hopefully will, be sufficient to balance the ship's weight, which is considerably less than it would be if it were made of solid steel. Therefore, it floats.

## Bernoulli's Equation

- Bernoulli's equation is an elegant mathematical formula that explains the relationship between pressure, fluid density, and fluid flow rate. It applies to everything from airplane wings to throwing a curveball to configuring the plumbing in your house.

- The flow of fluids can be steady or unsteady—that is, calm and smooth or choppy and turbulent. And just like water in a churning ocean or a still lake, air can also be smooth or turbulent. The physical properties of these different types of flow can be defined by the constancy or uniformity in the velocity of the fluid particles crossing a particular point in space.

- In addition to the steadiness or choppiness of fluid motion, different fluid substances can also be more or less compressed. Liquids and gases differ in their compressibility. The mass density of liquids is fairly constant despite pressure changes. That is, the density of water is the same at the bottom of the pool as it is at the surface—in other words, the number of molecules of water in a given amount of space or volume is the same whether you're at the surface or the depths.

- But the pressure is higher at the bottom, because there's more matter pushing down on an object at the bottom of a pool than at its surface. We feel this pressure in our eardrums as we dive to the bottom: Our eardrums can be squished by the pressure of the water, but the density, or the number of molecules per cube of water, remains the same.

- Gases are different, however. They can be compressed—that is, one can pack more and more molecules into an area. Tanks of oxygen are compressed in this way.

- Fluids can differ in terms of how they move and how much they can be compressed. They can also differ in terms of how they flow. Viscosity refers to how easily a fluid flows. We can think of viscosity like a measure of thickness. It's obvious that honey is thicker than water, and it flows more slowly and less easily than water, so honey has a higher viscosity than water.

- To understand fluid mechanics, we need to imagine what it would be like if we had an ideal fluid—an incompressible, nonviscous

fluid. It doesn't exist in the real world, but it's a useful construct for thought experiments.

- In a flowing material, we can appreciate that all of those molecules in space, which make up the density, are moving and therefore creating a mass flow rate. Water through a hose has a mass flow rate, which we observe as the water molecules streaming out the open end over time.

- Faster-flowing water has a higher mass flow rate than does slowly flowing water. Assuming a closed system—that is, no other forces act on it—the mass flow rate at the entry point must equal the mass flow rate at the exit point.

- Assuming that there are no holes in a hose (that is, it's a closed system), the flow rate of the water entering the hose must be the same as the flow rate of the water that exits the hose.

- This is known as the equation of continuity, and it explains why it helps to put your thumb over the end of a running garden hose to increase the speed of the spray of the water when you water your lawn. Your thumb partially obstructs the end of the hose, forcing the water to exit more quickly in order to maintain the same mass flow rate across the entry and exit points. The equation of continuity explains that the water seeks to conserve the amount of flow, so it must do so by forcing more water out a smaller hole.

- We now have what we need to understand Bernoulli's equation, a fundamental concept in fluid dynamics. Bernoulli's equation is named after 18th-century physicist Daniel Bernoulli, and it describes the relationship between pressure and two other factors, speed and elevation, for steady flow of an ideal fluid.

- For example, his equation predicts that for a flowing fluid at a constant elevation, such as water through a horizontal pipe, the fluid pressure will decrease as the flow rate increases. His equation also

predicts that for flowing fluid at a constant speed, the fluid pressure will decrease as elevation increases.

- What would make the flow rate increase under horizontal conditions? First, recall that pressure is the force exerted per unit area. As the mass flow rate increases, pressure decreases because those fluid molecules are less able to exert their force perpendicularly onto their surroundings. They're moving too fast to push on the walls of the pipe or whatever holds them.

- Think again about the equation of continuity: If we connect a large-diameter pipe to a small-diameter pipe and move water through it in a horizontal position, we would find that the flow rate increases and fluid pressure decreases as the water moves from the large- to small-diameter pipes. This is counterintuitive: You might think that as flow rate increases, so does pressure, but the opposite is true. This feature of fluid motion is what allows airplanes to fly.

- Recall that increasing the depth within a static fluid increases the pressure. Bernoulli's equation accounts for this as well within its description of fluids in motion. Specifically, Bernoulli noted that as the elevation of fluid increased, the pressure decreased. So, if you imagine a pipe of constant diameter with an uphill trajectory, a pressure gauge would read a lower pressure at an uphill point than at a downhill point for fluid moving at a constant flow rate.

- So, Bernoulli's equation describes the relationship of fluid pressure with fluid flow rate and with fluid elevation or depth. At a constant elevation, as the flow rate increases, pressure decreases. At a constant flow rate, as the elevation increases, pressure decreases.

- The physics of airplanes depends on Bernoulli's equation. An airplane wing is shaped such that airflow, the fluid in this case, moves faster over the wing than it does under the wing.

- The faster-moving air over the wing, in accordance with Bernoulli, will result in lower fluid pressure (air pressure) than under the wing,

resulting in an imbalance between the two sides of the wing. The air pressure under the wing will exert a greater force under the wing, producing a net upward lift.

- Extend the area of the wings enough, and generate enough airflow over the wings, and the resulting lift will be sufficient to pull a massive aircraft into the sky. Lift is analogous to buoyant force, which, when large enough, is sufficient to overcome the mass of an object so that it floats, so lift needs to exceed the mass of the plane to allow it to take off and stay airborne.

## Suggested Reading

Cengel and Cimbala, *Fluid Mechanics*.

Langeweische, *Stick and Rudder*.

## Questions to Consider

1. Why does turbulence on a plane feel so different from driving along a bumpy road?

2. What are some ways in which we intuitively apply the laws of fluid mechanics in our daily lives (examples similar to restricting the flow of a garden hose to generate more pressure)?

# Navigation and Propulsion in Fluids

## Lecture 16

When the object we're interested in—for example, an airplane—is moving through a fluid—the air around it—how can it best be powered? And what are the unique challenges of navigation in fluids? This is where thermodynamics meets fluid mechanics and concepts such as thrust and drag come into play. In this lecture, you will explore the nature of these concepts and what roles they play in moving us around the Earth.

**Thrust and Drag**

- Thrust is the forward force that pushes or pulls an object through a fluid. Modern airliners achieve thrust through the use of jet engines, which force air to stream through their cylindrical design. But thrust can also be generated by other means.

- Drag is an opposing force to thrust: It is the friction that opposes the motion of an object through fluid. In the case of objects moving through air, it is sometimes referred to as air resistance or aerodynamic drag.

- As thrust increases, so does drag, because they are opposing forces. So, an airplane needs thrust to exceed drag in order to accelerate in a forward direction and generate lift.

- Drag depends on a few factors. One factor is speed: As speed increases, so does drag. More air molecules will strike against the moving object per unit time, so the force of those air molecules will contribute to the drag that opposes the thrust.

- Two other factors are the area or size of the moving object and the density of the air. Larger, more voluminous objects generate more drag than do smaller objects, because the larger object gets in the way of more fluid molecules than a smaller object, hence the sleek design of cars and boats to reduce drag. Similarly, an object moving

through a high-mass-density fluid is subjected to more drag than is an object moving through a low-mass-density fluid.

- The three types of drag are form drag, skin friction drag, and lift-induced drag. Form drag is the drag created by the general size and shape of the object. Skin friction is the drag created by the fluid molecules in contact with the surface of the object. Lift-induced drag is a phenomenon resulting from the redirection of the air at the rear of a wing.

- One mathematical concept that ties together these various forms of drag is the drag coefficient, which incorporates the main forms of drag into one simple number. The drag coefficient includes in its denominator the mass density of the fluid, the speed of the object, and the area of a reference surface of interest, such as the area of the front of the wings, the bow of a ship, or the front of a car.

- All of the factors that we've discussed influence the amount of drag that opposes the thrust that is generating lift to get a plane off the ground. The drag coefficient is never a perfect description, but it can be used to compare the drag generated by a variety of different objects, everything from planes and cars to bicycles and buildings. In general, the higher the drag coefficient, the more drag the object creates.

**Turbulence**

- We've been considering how an airplane flies under near-ideal conditions, but the real world is more complex. There are weather patterns to consider—geological features to obstruct wind currents, which can lead to turbulence. Experiencing turbulence when flying gives us a direct sensation of the chaotic fluid dynamics around us.

- The word "turbulence" was used in the first lecture on fluid mechanics when different types of fluid flow movements were described. There is steady flow, or the uniform movement of fluid particles at constant rate and direction, and then there is turbulent flow, or the erratic and irregular movement of fluid particles.

- These terms apply to the air through which planes fly, too, so when we think of airplane turbulence, we're referring to the conditions through which the plane, or the wings specifically, are passing: erratic or changing air currents.

- As a plane flies, the airflow rates under and over the wings, if altered in variable ways, will produce subtle variations in the lift and drag. "Rough air," as it is called, commonly occurs as a result of changing air current and weather patterns, like storms, which have updrafts and downdrafts that can act on the plane and push it up and down, resulting in a bumpy ride.

- Pilots generally steer clear of storms, by flying over or around them. But geological features can disrupt airflow as well, creating turbulent conditions. For example, air sloping up and over the Rocky Mountains creates "mountain waves," or turbulent flow much like water breaking on a pier or coral reef.

- The global jet streams, the continuous airflow created by tropical and polar air currents, have an inherent variability as well that contributes to turbulent air. All told, these multiple factors give navigating in midair an inevitable factor of unpredictability.

- The next time you're flying, keep in mind two additional points. First, airplane design and construction can withstand an extraordinary range of stresses on the plane's body, far more than what it is subjected to during typical turbulence, so structurally, the plane is safe.

- Second, the statistics tell of an exceedingly small risk of bodily harm due to turbulence. Of the tens of millions of air passengers that fly in the United States every year, there are roughly 50 turbulence-related injuries, usually flight attendants or passengers caught without seat belts on. That's 0.000001% of passengers—an infinitesimally small incidence rate.

**Applications of Fluid Mechanics**

- What makes fluid mechanics such a fascinating area of study is that the principles apply to so many aspects of our daily lives, including acoustics, hydraulics, and plasma physics.

- We associate acoustics with hearing, sound, and sound quality, and that's largely what acoustics is all about, but technically, acoustics is the study of mechanical waves in any medium—gas, liquid, or solid. Because fluid mechanics deals with the properties of materials that flow, acoustics is considered a branch of fluid mechanics.

- Sound is the spectrum of pressure waves, waves that have both amplitude and frequency, which propagate through compressible fluids such as air or water. Our sense of hearing is possible because our inner ears are equipped to pick up small oscillations in air pressure that we then perceive as sound.

- Underwater acoustics follows the same basic principles as airborne acoustics, but of course, the properties of the fluid through which the waves are propagating—the water—come into play.

- One very important difference between air and water is compressibility and its inverse, elasticity. In general, water is rather incompressible, whereas most gases, including air, are very compressible. It turns out that the speed of sound is highly dependent on the compressibility and density of the fluid through which the sound waves are propagating. The less compressible a substance, the faster sound travels through it.

- In general, the speed of sound is higher in solids than in liquids, and it's higher in liquids than in gases. This is because compressibility is just the inverse of elasticity. Liquids are more elastic than gases because when molecules in those fluids are moved from their starting position, there is a restoring force to get them back into position. There is almost no restoring force in gases like air.

- Although water is denser than air, it is also much less compressible, so sound travels about four times faster in water than it does in air. This relationship between sound travel and the compressibility of the medium or fluid holds true for other materials as well. For example, sound travels faster when propagating through the steel rails of train tracks than when propagating through air. We can hear the train coming along the track if we put our ear next to the track well before we can hear its engine in the air.

- The incompressibility of water and air come into play in other applications of fluid mechanics: hydraulics and pneumatics. Hydraulics is the study and application of pressurized liquids, generally with the goal of transferring energy, whereas pneumatics uses pressurized gas to move things.

- Hydraulics, as applied to water, has been harnessed by human societies since the time of the ancient Romans, and we still utilize hydraulics today in a variety of situations. Central to the applications of hydraulics is Pascal's principle, the notion that pressure changes are distributed uniformly throughout a fluid in a closed system.

- This is a very useful finding because force is equal to pressure multiplied by area. So, if one increases the area and applies the same pressure, the force generated also increases. The field of pneumatics applies essentially the same principles, but the medium is gas rather than liquid.

- While liquids and gases often act in similar ways, there is another medium that is akin to a combination of the two that is slowly and quietly revolutionizing fluid mechanics: plasma. At the frontier of fluid mechanics is an emerging subfield that has already yielded some life-changing applications: plasma physics.

- Plasma is a transitional state that occurs when a gas is exposed to very, very high temperatures, which causes the atoms within the gas to ionize. Ionization means that the electrons are separated from their protons. When this happens, the result is a mass of charged

particles. And, even though plasma is neither a liquid nor a gas, many of the principles of fluid mechanics still apply.

- Plasma is essentially net neutral in its charge as a medium, because the positively and negatively charged particles are present but unbound. In fact, a recombination will spontaneously occur when the conditions for ionization do not persist. For the majority of applications of plasma physics, we can treat plasma as a fluid of charged particles, in which electromagnetic forces are accounted for. That's why fluid mechanics still applies.

- We come across plasma in several ways in our world—for example, the extreme heat of a lightning bolt temporarily ionizes the air around it, producing plasma—but plasma is actually the dominant material state in the universe, both in terms of mass and volume. The Sun and the twinkling stars in our night sky are all composed of plasma. Because fluid mechanics applies to all of it, it's among the most important fields of science.

- Because plasma is electrostatically charged due to the separation of positively charged protons and negatively charged electrons, it has certain fluid properties that make it unique among the other three material states: gas, fluid, and solid.

## Suggested Reading


Anderson, *Fundamentals of Aerodynamics*.

Drummond, *Plasma Physics*.


## Questions to Consider

1. Of all the vehicles, bicycles, and other forms of transportation that you use, which is the most aerodynamic? Why?

2. Why are plasma televisions such an improvement over their predecessors?

# The Big Bang That Didn't

## Lecture 17

Although it accounts for most if not all of the available evidence collected thus far, the big bang theory of the origin of the universe is still a bit bizarre: The idea that everything started from a single point and that our universe continues to expand is baffling. However, there's a beautiful elegance to the theory that has captured the imaginations of our era's greatest thinkers. In this lecture, you will evaluate this theory by considering the evidence, and you will ponder the mind-boggling implications that must follow if the theory is, indeed, fact.

### The Big Bang

- The main concept behind the big bang theory is that the entire universe began with a bang, an explosion of matter that created all that we know. But this view of the big bang is actually incorrect in many ways.

- For one thing, nothing actually banged. For another, matter didn't even exist when it happened and, even today, seems to be only a miniscule portion of what the big bang produced. Furthermore, the big bang theory actually explains what happened *just after* time began and the universe was an infinitesimal speck of energy.

- First, despite its name, the big bang was not an explosion—nothing blew up. Instead, the tiny speck of the beginning of the universe began its colossal expansion 13.7 billion years ago, in an instant. From a tiny dense point, all of space and time expanded outward, initially at a very fast pace and eventually slowing down.

- The second big idea is that at the beginning of the big bang, there was no matter; the speck that eventually became everything in our universe was infinitely dense and extremely hot. As the universe expanded, it lost density and heat, and eventually, matter began to form, and through radiation, it began to lose energy.


**Although they are not as popular as the big bang theory, there are other theories that attempt to explain how the universe began.**

- The third idea to keep in mind is what the theory describes: what happened after the expansion began. Not only all space but also all time began with the expansion. What could have come before, if there was a before, is still in the realm of speculation.

## Evidence for the Big Bang

- How do we evaluate a theory that provides an explanation for something that may or may not have happened 13.7 billion years ago? To put that number in perspective, the entire history of human civilization covers about 10,000 years.

- To evaluate the big bang theory, we need to delve into two fields of physics: astronomy, or the observations of stars and celestial bodies, and cosmology, or the study of the astrophysical properties of the universe.

- The origin of the fundamental idea that the universe expanded from a single point and continues to expand today is attributed to Georges

Lemaitre, who, in the 1920s, first devised a model of an expanding universe and proposed that the expansion would have an effect on the wavelength of electromagnetic energy traveling through space. Specifically, he predicted that light emitted from distant stars and galaxies would look more reddish than light from closer stars.

- As stars in other galaxies move farther away from us, there is a shift in the signature of hydrogen as measured by spectroscopy: The light coming from the faraway hydrogen atoms is shifted farther down the wavelength spectrum. It's called redshift, and it is evidence that galaxies are receding away from us.

- This redshift effect is also called the relativistic Doppler effect. Think of a siren from an ambulance: Coming toward us, it sounds high because the frequency of the sound wave that we perceive is higher. In the instant that it passes us, we hear it at the same frequency as it is emitted, and as it drives away, the siren sounds lower, as the frequency of the sound wave decreases. As galaxies recede from us, just like an ambulance siren, the frequency of the wavelength of light that we observe decreases—and becomes more red, hence the redshift.

- U.S. astronomer Edwin Hubble had already published his observation that the starlight from more distant galaxies reddens more than closer ones, and these observations were confirmed in 1929.

- So, Lemaitre extrapolated from this information that if the universe is expanding today, it must have started at some point in the distant past with a size of zero, or very, very small. Lemaitre called his hypothesis the primeval atom, and while he was eventually proved right, even Albert Einstein found his notion a bit ridiculous at first.

- The next major point came from Russian theoretical physicist George Gamow, who wrote a paper along with Ralph Alpher on big bang nucleosynthesis, which explains how the current levels of hydrogen and helium in the universe were created by reactions during the big bang.

- The point of his paper was the idea that if the early universe was hot enough, it would act as a pressure cooker: At these high temperatures, basic particles like electrons, protons, and neutrons would be moving around too quickly to combine with each other. But, as the universe expanded, and the space cooled, the neutrons and protons could come together to form nuclei of elements.

- This idea was particularly useful in describing the formation of the lightest elements, hydrogen and helium, and provided an account for their abundance in our universe. Our universe is around 74% hydrogen and 25% helium, with all the other elements put together representing a tiny percentage of matter. The paper written by Alpher and Gamow predicted precisely this state of affairs.

- The calculations that Alpher conducted didn't work so well for the heavier elements, however, and British mathematician and astronomer Fred Hoyle set out to disprove the big bang theory by explaining their formation in the stars in his famous 1957 paper.

- But Alpher and Gamow also realized that the big bang would leave another signature: a thermal radiation that should be found in trace amounts everywhere in the universe. And, in fact, cosmic microwave background radiation, or relic radiation, has been discovered in distributions that fit that description. The big bang is still considered the best explanation for this relic radiation pattern and its fluctuations.

- The third test of the big bang theory came in the form of the shape of the contents of the universe. The big bang theory makes certain predictions with respect to how galaxies and quasars are distributed in space and how they are shaped.

- For example, in the first few moments after the big bang, the theory predicts that there was a period of rapid expansion, or inflation. This inflation would have been enabled, in part, by the fact that the universe at that point would have been made up of an unstable form of energy that we still don't understand.

- This energy would cause the matter in the universe, galaxies, and so on to be unevenly dispersed. In fact, the very sort of pattern that the theory predicts is visible in snapshots taken by a satellite of background radiation from about 400,000 years after the big bang.

- Yet another source of evidence comes from quasars and other galaxies. Quasars are the most distant visible objects; they are massive black holes surrounded by a flat disk-shaped gas called an accretion disk. They are extremely bright, but they look red to us because they are so far away and are receding from us.

- The first quasars and galaxies were formed about 1 billion years after the big bang. And we can observe them in different stages of their evolution. Because light from faraway galaxies must travel all the way to Earth to be observed, what we see of them today represents what they were a long time ago. The farther away the galaxy, the older the snapshot that we glimpse.

- So, we can compare observations from nearer galaxies with those from distant ones and get a sense of how galaxies evolved over time. These observations agree with simulations that are generated by models of the big bang. They also refute the predictions of the biggest rival to the big bang theory—the steady state theory championed by Fred Hoyle, among others.

- The steady state theory posits that matter is eternal and that new matter is created as the universe expands. There is no need for a beginning and an end in this theory. But, given the accumulation of evidence in favor of the big bang, this theory is now considered obsolete. It seems that there was a finite beginning, and there probably will, eventually, be an end. But the end of the universe is a long way away.

- In 2011, more evidence for the big bang was discovered in very distant quasars. By analyzing the absorption lines of the spectra of these quasars, astronomers found clouds of primordial gas: Unlike all the other gas clouds observed, these clouds contained only two

elements, and no trace of the heavy elements made by the stars. There was only hydrogen and deuterium, also known as heavy hydrogen because it's an isotope of hydrogen.

- These observations prove that at some point during the life of the universe, matter consisted almost exclusively of hydrogen—a prediction stemming from the big bang. And there are still other lines of evidence supporting the theory, primarily related to the age of the universe and its contents.

**Three Additional Theories**

- The veracity and explanatory power of the big bang theory is great, so if we take it for granted that the theory is correct, what are its implications? Where do we go from here? Unfortunately, there are three main options going forward—the big crunch, the big freeze, and the big rip—and none of them are desirable.

- The critical density of the universe is the value at which the universe is in balance and expansion stops. We can compare the actual density with this critical value and get a ratio that tells us whether the universe is collapsing, expanding, or flat.

- If the mass density were to continue to increase, then eventually the universe would max out in terms of size and begin to collapse onto itself. It would become more dense and hot once again, and the big bang would happen in reverse. This is the idea known as the big crunch.

- If the universe didn't reach the critical mass density, then it would continue to expand and cool, and stars would burn out, leaving black holes in their wake. These large masses of matter would eventually collide and form even larger black holes, and the temperature of the universe would reach absolute zero. This is the big freeze.

- However, what neither of these two scenarios take into account is dark energy, a hypothetical form of energy that accelerates the expansion of the universe. Indeed, based on our observations, the

universe is expanding at an accelerated rate, so scientists have proposed that dark energy exists to explain this phenomenon. There is no direct evidence for dark energy, but it does seem to provide the missing link in explanations of other observations of the cosmos.

- In 2003, the big rip was proposed: a hypothesis that predicts that the matter of the universe, stars, galaxies—all atoms—will be torn apart by the expansion of the universe at a specific time in the future. At some finite point in time, the expansion will become infinite.

- Everything we think we know about dark energy might turn out to be false once we can finally make some direct observations. At that time, all bets are off as to how these predictions will pan out.

## Suggested Reading


Hawking, *A Brief History of Time*.

May, Moore, and Lintott, *Bang!*


## Questions to Consider

**1.** How does our understanding of the big bang and the start of the universe relate to your sense of the meaning of life?

**2.** Why are we so curious about how time and our universe began?

# The Four Forces of Nature

## Lecture 18

This lecture explores the four fundamental forces of nature—the electromagnetic force, the strong nuclear force, the weak nuclear force, and gravity—which we think have guided the formation, expansion, and essence of the universe since that first trillionth of a trillionth of a trillionth of a second. Understanding what we have learned about these forces, and their interactions, is one of the keys to understanding our world and predicting how matter in the world will behave under different conditions.

**The Four Forces**

- At the big bang, there was just one unifying force—or so research suggests. But as the universe began to expand and cool, four fundamental forces that still govern the universe today began to emerge. How these forces were unified in the beginning remains one of the great mysteries of physics.

- Perhaps the most compelling argument is that at the start of the big bang, when the precursors of atoms and their particles were so hot and squished together that their boundaries were undefined, the forces, too, were unified in one pot of universal soup.

- But as the universe expanded and the ambient temperature cooled, the strongest of the four fundamental forces, which is called simply the strong force, separated from the others. Shortly thereafter, two more forces, which we call the electromagnetic and weak forces, began to exert their unique powers. How the last force, gravity, came to be in those first few moments remains unknown.

- The strong force holds the nucleus of the atom together. And splitting the nucleus is what causes nuclear fission, the energy source that powers atomic bombs. The weak force is what's responsible for radioactivity, the force that governs the change of

a neutron into a proton and an electron, which is then ejected from the nucleus during radioactive decay.

- What repels that electron from the nucleus is the electromagnetic force, in which opposite charges attract and like charges repel. Charge describes the behavior of subatomic particles: Negatively charged particles like electrons are attracted to positively charged protons and are repelled by other electrons. We don't really know how a particle gets its charge in the first place; it's a fundamental property of the particle. But we can predict how it will behave—at least most of the time.

- Finally, there is the mysterious force of gravity, which is arguably the most perplexing force because, on a large scale, it's the strongest force. It's what is keeping our world orbiting the Sun. But, on a smaller scale, it is trumped by all of the other forces.

**The Electromagnetic Force**

- Electromagnetism is a fundamental part of atoms and their structure. There are two possible electric charges: positive and negative. Why aren't there other types of electric charges? We can barely imagine what these other charges could be, and yet there are different types of spin, for example, in subatomic particles. They can show a half-integer spin, for example. It's not just clockwise and counterclockwise; charged particles could also have a two-thirds charge or a half charge.

- Instead, protons have a positive charge, and electrons have an exactly symmetrical negative charge, such that the two cancel each other out entirely. In nature, in fact, the vast majority of atoms are neutral: They house an equal number of protons and electrons, though different parts of atoms might have slightly more or less charge, depending on exactly where the electrons are in their orbits.

- Just like light, then, electric charge is said to be quantized: It comes in indivisible magnitudes of charge, like photons of light. There's no such thing as half of a charge, but an ion or a molecule with

extra charged particles can show a charge of, for example, 3+ or 2−, depending on the imbalance between electrons and protons.

- The electromagnetic force incorporates both electric charges and magnetic fields. For a long time, electricity and magnetism were considered to be distinct forces, as astute observers discovered rocks that seemed to attract certain materials and the effects of lightning and other natural forms of electricity.

- However, in the 19th century, there was a surge in interest and excitement in the field of electrical science. This fruitful period of scientific discovery resulted, by the end of the 19th century, in the harnessing of electricity and the sweeping changes that bringing it into homes caused.

- In 1873, James Clerk Maxwell published a series of equations that show how electricity and magnetism are related. Specifically, he suggested that electricity and magnetism are, in fact, one force—the electromagnetic force.

- In addition to affecting relatively large objects, the electromagnetic force acts on subatomic particles. The electromagnetic force acts both on very short and relatively long distances, and the range of the force is infinite, like gravity. But unlike gravity, there isn't a great difference between large and subatomic versions of the electromagnetic force.

**The Strong and Weak Nuclear Forces**

- There was one mystery that understanding the electromagnetic force highlighted in terms of atomic structure—namely, how does the nucleus hold itself together if it's made up of positively charged particles? Like repels like, after all.

- This observation led to the discovery of the strong nuclear force, a force that we've harnessed to create and destroy in unfathomable proportions. The strong nuclear force is what holds the nucleus

together, and the first person to model it was Japanese scientist Yukawa Hideki in 1935.

- Hideki suggested that when protons and neutrons interact with each other, they exchange a particle called a meson, which transmits the force. It wasn't until the 1950s, when the first particle accelerators were built, that this particle was finally observed.

- Working with particle accelerators, physicists also found that the protons and neutrons were in fact made up of even smaller quarks, and the strong force then holds the quarks together, carried by gluons, which then form the nucleus.

- Like electric charge, gluons also carry a charge, but there are three possibilities, instead of just two. The charges that gluons can carry are called color charges, and each results is a different type of force.

- The strong nuclear force is 100 times stronger than the electromagnetic force at the subatomic level. It is the binding energy in the nucleus that is exploited using nuclear physics to create the intense reactions and energy release of nuclear power and the atomic bomb.

- While the strong nuclear force keeps the nucleus together, there is another force that governs the behavior of subatomic particles, and that's the weak nuclear force. As its name suggests, it is weaker than the strong force, and it's responsible for radioactive decay.

- The weak force changes quarks from one type into another by exchanging force-carrying particles called bosons, as was discovered in 1983. Over time, most quarks, and a related set of particles called leptons, eventually decay via the weak nuclear force. This decay is what makes radiocarbon dating possible.

**Gravity**

- Gravity, unlike electromagnetism, is not picky—it's the force that pulls together all physical bodies and is directly proportional to

mass, so the more massive an object, the greater its gravitational pull.

- However, contrary to our own personal experience, gravity is the weakest of the four forces, at small scales. It feels to us as though it's quite strong, because it keeps us attached to the ground, and defying gravity is difficult. But when you compare its strength with the other force that we can see—electromagnetism—you get a sense for how weak it really is.

- When at least one of the two objects we're considering is small, gravity is weak. But when both objects are massive—like the Earth and the Sun—gravity is in fact the strongest force, overcoming all of the others.

- On Earth, and elsewhere, the way in which gravity affects medium-sized objects (larger than atoms but smaller than planets) is characterized by Newton's laws, at least for objects traveling at low speeds.

- Newton's law of universal gravitation breaks down gravity into the following two rules: Every mass-bearing particle attracts every other with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. In a nutshell, things with mass are attracted to each other with a force that increases as their masses increase and the distances between them decrease. Bigger and closer means greater gravitational pull.

- However, Newton's laws break down at both extremes: for very small things, that travel fast, and very large things, like Mercury's orbit around the Sun. Mercury is the planet closest to the Sun, and the eccentricity of its orbit puzzled astronomers for centuries. Newton's laws could not account for it.

- In his general theory of relativity, Einstein combined his ideas on special relativity with Newton's law of universal gravitation, and

general relativity, as it's now called, is still the current description of gravitation in physics.

- In its simplest form, special relatively describes the fact that you can't really tell if you are moving unless you look at another object—that is, movement is relative. In addition to this idea, Einstein's theory of relativity had several other corollaries. The theory of relativity demonstrates that concepts previously considered separable are, in fact, intrinsically linked, like space and time, and, perhaps most famously, energy and mass.

- Arguably the most famous equation in all of science, let alone physics, is $E = mc^2$, where $E$ is energy, $m$ is mass, and $c$ is the speed of light. Einstein's insight was that energy and mass are like convertible currencies—we can change one into the other. But the exchange rate is always fixed; it's the speed of light squared. And because the speed of light is a big number, the energy generated by even a small increase in mass is very large.

- Instead of thinking of gravity as a force, Einstein suggested that we think of it in terms of the curvature of space-time, which is a notion in physics that puts both space and time on a single continuum—so that we can talk about its shape as one thing.

- In this model, gravity creates curves in space-time, by bending it in specific ways. The best way to imagine this phenomenon is to think of what would happen if you put a bowling ball on a trampoline: The trampoline surface would bend to accommodate the weight of the ball.

© Monkey Business/Thinkstock.

**The surface of a trampoline bending as a result of weight being dropped on top of it can be used to conceptualize the complex concept of gravity creating curves in space-time.**

- Then, if you think of the ball as the Sun, and a marble as the Earth, what would happen if you dropped the marble on the trampoline? It would be attracted to the ball, and its motion would be affected by the curvature of space-time caused by the bowling ball.

- If the marble is traveling very quickly, it can escape the depression caused by the ball, but its motion will still very likely be affected by it. In this way, Einstein conceptualized gravity not as a force per se, but as a distortion of space and time caused by the mass of an object. The larger the object, the greater the distortion.

## Suggested Reading

Cox and Forshaw, *Why Does* E = mc$^2$ *and Why Should We Care?*

Tyson, *Death by Black Hole and Other Cosmic Quandaries*.

## Questions to Consider

**1.** What would our lives be like if gravity were a weaker or a stronger force?

**2.** Is it possible that there is a fifth force that we have yet to discover? If so, what might it be?

# The Elements of Everything

## Lecture 19

An atom is a dense nucleus of positively charged protons and neutral neutrons, surrounded by a cloud of negatively charged electrons. Electrons are bound to the nucleus by the electromagnetic force; they are attracted to its positive core. We classify an atom based on the number of protons in its core, although the chemical properties of an atom are defined by the number of electrons that surround the core. In this lecture, you will learn about the elements that make up matter and the elegant periodicity with which their properties can be predicted from their atomic makeup.

**The Periodic Table of the Elements**

- Chemistry is the study of the composition, properties, and behavior of matter. All of the known elements of matter, pure chemical substances made up of only one type of atom, are represented in an orderly fashion in the periodic table.

- When Russian chemist Dmitry Mendeleev published the first periodic table, the idea that there is an order and periodicity to the elements caused a paradigm shift and an explosion in the study of the chemical properties of matter.

- The table is an 18×7 arrangement of rows and columns, plus a double row on the bottom that can be further divided into blocks and columns. The current table is read from left to right in increasing order of atomic number, which is the number of protons in the core of the nucleus of an atom. But the number of neutrons can vary, creating different isotopes of the same element.

- The periodic table starts with hydrogen, which is the simplest of the elements—being made up of only one proton—and the first element formed after the big bang. In fact, the most common isotope of hydrogen just has one proton, and no neutrons. It's also the most

common element in our universe, accounting for roughly 74% of the atomic mass (of all the known atomic matter) in the universe.

- After hydrogen comes helium, the next lightest and most abundant atom. With an atomic number of 2, it accounts for 23% to 25% of the known mass of the observable universe. Helium makes up most of the mass of the Sun and other stars. On Earth, it's actually pretty rare.

- Both hydrogen and helium are very light but have very different properties. They have their own row on the periodic table. There is some disagreement, however, as to their placement there; they don't really behave like the other elements in their columns.

- The rows of the periodic table indicate how many electron shells a given element has, and they are called periods—hence the name "periodic table." The columns are called groups. The electron shells can be thought of as the orbital routes that electrons can travel around a nucleus. They represent levels of energy excitation in the electrons.

- The first row of the table contains only helium and hydrogen, because they are the only elements that have only 1 electron shell. Each shell can only hold a limited number of electrons, with the first shell holding just 2.

- As you move from left to right along the period, the radius of each atom *decreases*—this is because each increase in atomic number, or added proton, draws the electrons closer to the nucleus. Also, the more tightly the electrons are bound to the nucleus, the harder it is to remove one; therefore, the electron affinity increases as you go from left to right.

- This periodicity of the table allows us to explain why elements behave the way they do and predict the existence and properties of elements we haven't yet seen in nature or synthesized in the lab.

- In the early models of atoms, electrons were thought to orbit around the nucleus in their shells—with every electron in each shell having

# The Periodic Table of the Elements

| Period | Group 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 H 1.008 | | | | | | | | | | | | | | | | | 2 He 4.003 |
| 2 | 3 Li 6.941 | 4 Be 9.012 | | | | | | | | | | | 5 B 10.81 | 6 C 12.01 | 7 N 14.01 | 8 O 16 | 9 F 19 | 10 Ne 20.18 |
| 3 | 11 Na 22.99 | 12 Mg 24.31 | | | | | | | | | | | 13 Al 26.98 | 14 Si 28.09 | 15 P 30.97 | 16 S 32.07 | 17 Cl 35.45 | 18 Ar 39.95 |
| 4 | 19 K 39.10 | 20 Ca 40.08 | 21 Sc 44.96 | 22 Ti 47.88 | 23 V 50.94 | 24 Cr 52 | 25 Mn 54.94 | 26 Fe 55.85 | 27 Co 58.47 | 28 Ni 58.69 | 29 Cu 63.55 | 30 Zn 65.39 | 31 Ga 69.72 | 32 Ge 72.59 | 33 As 74.92 | 34 Se 78.96 | 35 Br 79.9 | 36 Kr 83.8 |
| 5 | 37 Rb 85.47 | 38 Sr 87.62 | 39 Y 88.91 | 40 Zr 91.22 | 41 Nb 92.91 | 42 Mo 95.94 | 43 Tc (98) | 44 Ru 101.1 | 45 Rh 102.9 | 46 Pd 106.4 | 47 Ag 107.9 | 48 Cd 112.4 | 49 In 114.8 | 50 Sn 118.7 | 51 Sb 121.8 | 52 Te 127.6 | 53 I 126.9 | 54 Xe 131.3 |
| 6 | 55 Cs 132.9 | 56 Ba 137.3 | 57 La 138.9 | 72 Hf 178.5 | 73 Ta 180.9 | 74 W 183.9 | 75 Re 186.2 | 76 Os 190.2 | 77 Ir 192.2 | 78 Pt 195.1 | 79 Au 197 | 80 Hg 200.5 | 81 Tl 204.4 | 82 Pb 207.2 | 83 Bi 209 | 84 Po (210) | 85 At (210) | 86 Rn (222) |
| 7 | 87 Fr (223) | 88 Ra (226) | 89 Ac (227) | 104 Rf (257) | 105 Db (260) | 106 Sg (263) | 107 Bh (262) | 108 Hs (265) | 109 Mt (266) | 110 Ds (271) | 111 Rq (272) | 112 Uub (285) | 113 Uut (284) | 114 Uuq (289) | 115 Uup (288) | 116 Uuh (292) | 117 Uus 0 | 118 Uuo 0 |
| | | | 6 | 58 Ce 140.1 | 59 Pr 140.9 | 60 Nd 144.2 | 61 Pm (147) | 62 Sm 150.4 | 63 Eu 152 | 64 Gd 157.3 | 65 Tb 158.9 | 66 Dy 162.5 | 67 Ho 164.9 | 68 Er 167.3 | 69 Tm 168.9 | 70 Yb 173 | 71 Lu 175 | |
| | | | 7 | 90 Th 232 | 91 Pa (231) | 92 U (238) | 93 Np (237) | 94 Pu (242) | 95 Am (243) | 96 Cm (247) | 97 Bk (247) | 98 Cf (249) | 99 Es (254) | 100 Fm (253) | 101 Md (256) | 102 No (254) | 103 Lr (257) | |

the same energy level. That is, picking off any electron from a shell takes the same amount of energy.

- With advances in physics and chemistry, however, scientists realized that electrons can take several different paths around the nucleus, regardless of their energy level, and that these paths take predictable, repeating shapes. Each orbital shape, then, was defined as a subshell. So, you can have more than one type of shape, or subshell, at the same energy level or shell.

- The orbitals follow a predictable pattern: The first 2 electrons orbit the nucleus like a sphere, the next 6 electrons take routes that look like dumbbells, and the next 10 take more complicated pear-shaped routes.

- Each row in the periodic table defines a different number of shells or energy levels, and each of these different routes, or subshells, defines a different block. One block is called the s-block and represents the spherical route; the p-block is the ellipsoids, and the d-block is the pear-shaped lobes. Finally, the f-block has even more complicated orbital routes.

- The orbital patterns of the outermost electrons dictate the properties of the elements in those blocks. The periodic table is an elegant representation of one of Mother Nature's marvels: Each of the elements in a block is similar in terms of chemical makeup and behavior because of how electrons orbit around the nucleus.

**Subatomic Particles**

- Subatomic particles are particles inside the atom that are even smaller than the atom itself and obey the laws of quantum mechanics. A model of subatomic particles known as the standard model took shape in the 1970s. For some physicists, it's sometimes thought of as the core of the "theory of everything." There are still aspects of the universe that it doesn't cover, so it falls short, but it has huge explanatory power.

- The standard model predicts that there are elementary particles that cannot be further broken down: We used to think that atoms were elementary, and then that protons, neutrons, and electrons were elementary, but now we're quite confident that the elementary particles are fermions and bosons.

- Particles are categorized as either fermions or bosons based on their spin: Fermions have half-integer spin, and bosons have integer spin. Spin is the angular momentum of a particle—its rotation on its axis, sort of like a basketball spinning on your finger.

- Elementary particles, like different pitches in baseball, have individual spin trajectories. We classify curveballs and knuckleballs based on how the ball travels through the air. So, we classify fermions and bosons based on how the particle moves, too.

- Matter is made up of fermions, of which there are 12 kinds, or flavors. Force is carried by bosons, which can either be gauge bosons or the Higgs boson. There are also 2 types of fermions: elementary and composite. Elementary fermions consist of leptons, which include electrons and their buddies, and quarks. Composite fermions, which are also known as hadrons, include protons and neutrons, which are both made of various kinds of quarks.

- The 6 flavors of quarks are up, down, top, bottom, charm, and strange. Up and down quarks are the lightest, in terms of mass, and the other 4 types of quarks rapidly change into up and down quarks as particles decay. Once they change, they remain relatively stable. Quarks can't be directly observed and can only be found as part of composite fermions, or hadrons, like in protons and neutrons.

- The most famous lepton is the electron, whose activity affects much of our lives, from our biological functions to electricity to chemistry. Leptons also come in two forms: charged particles, like electrons, and neutral ones, called neutrinos. Charged particles interact with others, but neutrinos keep to themselves and, therefore, are rarely seen.

- Just like there are 6 flavors of quarks, there are also 6 flavors of leptons. These flavors can be further categorized into 3 types, called generations, with each generation composed of one charged lepton and one neutrino. The 3 generations are electron, muon, and tau—so, we have the electron and the electron neutrino, the muon and the muon neutrino, and the tau particle and the tau neutrino.

- Like charm, strange, top, and bottom quarks, the tau and muon leptons have a larger mass than electrons, so they rapidly decay into electrons and can only be produced in high-energy collisions.

**The Higgs Boson**

- There is one last set of subatomic particles to consider: the bosons, which are named after physicist Satyendra Nath Bose, who, along with Einstein, described the ways in which particles with integer spin values behave.

- What differentiates bosons and fermions is their spin: Fermions have a half-integer spin, while bosons have an integer spin. This means that the wave functions that describe the routes that fermions take as they travel through space are not symmetrical. Bosons, in contrast, follow a wave function that is symmetrical.

- Another difference is that although all fermions of the same flavor are the same—that is, electrons repel each other, so two electrons cannot share the same quantum state simultaneously—bosons can share quantum states. This feature, along with the fact that they have an integer spin, makes elementary bosons act like glue: They hold matter together.

- For decades, the standard model of physics has postulated that there are 4 gauge bosons: photons, which carry the electromagnetic force; W and Z bosons, which carry the weak interaction force; and gluons, which carry the strong interaction force.

- Perhaps the most famous boson is the long-elusive Higgs boson, which was the inspiration behind the Large Hadron Collider, a massive particle accelerator in Switzerland.

- The Higgs boson was conceived of in the 1960s to explain certain inadequacies of the standard model as it was then formulated. But first, physicists suggested that there is a Higgs field, existing throughout space, and it was thought to be responsible for the weak nuclear force and for how some very light particles like quarks and electrons get their mass.

- Then, the Higgs boson was believed to be a by-product of this field when the field was excited, and physicists believed that the existence of the field could be proven if the Higgs boson was observed.

- On July 4, 2012, the first evidence of the Higgs boson was observed, and it was confirmed in March of 2013. The observation of the Higgs boson gives physicists hope that we may soon gain a new understanding of how things with mass interact—knowledge that may change the world as we know it.

- There is one more subatomic particle that is still rumored to be out there, and that's the graviton, which is thought to be the force-carrier particle for gravity. Despite the fact that we all experience it every day, gravity remains outside the scope of the standard model. Einstein's theories provide an explanation for it, but the search for proof of those theories continues.

## Suggested Reading

Gray and Mann, *The Elements*.

Kean, *The Disappearing Spoon*.

## Questions to Consider

1. How would our world be different if atoms didn't follow a periodic and predictable structure?

2. What are some other ways by which we could format the periodic table of the elements?

# Looks like a Particle, Acts like a Wave

## Lecture 20

In this lecture, you will learn that atoms and their component particles don't always act the way that they're supposed to. In fact, the smallest bits of matter sometimes act more like waves than like particles. The study of the behavior of particles is called quantum mechanics. Understanding how particles behave—how they can be both particle-like and wavelike—has already yielded many fruitful applications, and we still don't even know exactly how this dual nature works.

**Understanding Light**

- One of the most important scientific discoveries in quantum mechanics, or any area, began with a controversy. In the 17th century, physicists were particularly interested in understanding the nature of light, and two great minds proposed opposing theories: Christiaan Huygens suggested that light is made up of waves, and Isaac Newton insisted that it consists of particles.

- In the early 1800s, Thomas Young conducted an experiment that involved a light source, a thin barrier with two vertical slits, and a photographic plate. If we place the light source in front of the barrier and point it toward the plate, as the light source sends out light, the barrier dictates where it goes—it must pass through the slits—and the photographic plate captures where it ends up. The brighter the image on the plate, the more light reaches that point.

- If we turn on our light source and cover one of the slits, then we see a single line of light on the photographic plate, corresponding to the remaining open slit. But when both slits are open, Newton's theory predicts that we'll see two vertical lines of light, each corresponding to one of the slits.

- But Huygens's theory makes a different prediction: If light acts as a wave, undulating up and down, then when it passes through the

slits and gets separated, the two waves will interact with each other. In places where the peaks of the waves hit each other, you'll see a greater concentration of light, and where the valleys overlap, you'll see less light. But there will also be times when a peak and a valley hit each other, cancelling each other out.

- So, if light travels like a wave, you'll see a different pattern on the photographic plate—it will look like a whole series of vertical lines, not just two, of different brightnesses: the brightest ones showing where two peaks coincided and the dark areas where they cancel each other out. We call this sequence of light and dark bands an interference pattern.

- When Young turned on his beam of light, he saw precisely this interference pattern, not the two simple lines that we would see if light behaved as a particle, so Newton's theory was proved wrong.

- Toward the end of the 19$^{th}$ century, science was in a quandary over the subject of energy. In 1900, scientist Max Planck made a guess that would eventually win him the Nobel Prize in 1918 and would question just how wavelike light really is.

- It seems that a bunch of physicists and mathematicians used their well-established formulae to try to calculate how much energy is inside an oven when it's heated to a certain temperature. They found that no matter what temperature was reached inside the oven, the total energy in the oven was calculated to be infinite.

- Of course, the group realized, the calculations must be wrong. But it wasn't until the year 1900, when Planck made a fortuitous guess as to why they were wrong, that the problem found a solution, and the field of quantum mechanics was born.

- Planck's guess was that energy doesn't operate on a continuum; it comes in units. There is some fundamental energy denomination, and then energy can increase only by multiples of that initial sum. No fractions are allowed.

- But how are the values of these denominations determined then? Planck suggested that the minimum energy that a wave can have is determined by its frequency—that is, how often the peaks and valleys alternate. This relationship is inversely correlated: The smaller the frequency or the longer the wavelength, the lower the minimum energy. This principle applies to light as well as to heat.

- Planck's discovery that the energy in waves comes in lumps, rather than continuously, won him the Nobel Prize in 1918, but he didn't tell us why energy behaved this way. It was Einstein who provided the "why," and he won his own Nobel Prize for it in 1921. Einstein figured out that energy can only be increased in certain increments—because the energy is triggered by the ejection of an electron from a metal, or other atom.

- Einstein's insight as to why energy is lumpy is based on his explanation of something called the photoelectric effect, which describes the observation that when you shine a light on certain metals, they eject electrons. What controls whether or not electrons are ejected is the wavelength of the light—not its intensity. Light of different wavelengths looks to us as though it has different colors, with red light having the longest wavelengths and violet light having the shortest.

- Einstein suggested that we should think of light as a stream of tiny packets—eventually called photons, each with an energy value that is proportional to the wavelength or frequency of the light wave. If the light beam's wavelength is too long, and its frequency therefore is too small, the photons in that beam won't have enough energy to knock out electrons.

- Einstein showed that Planck's guess that electromagnetic energy is lumpy results from the fact that waves are composed of little bundles, photons, or quanta of light. Just like water is made up of molecules, and therefore waves in the ocean can be broken down into particles, light is made up of photons. Newton and Huygens were both right.

**The Heisenberg Uncertainty Principle**

- Austrian physicist Erwin Schrödinger proposed a famous thought experiment to illustrate one of the tenets of quantum mechanics—that if you put a cat into a box, along with some radioactive substance, the cat is both alive and dead at the same time.

- Schrödinger also suggested that the wavelike trajectory of electrons should be thought of as a smeared out electron, like blueberry jam smeared on toast. However, this description wasn't particularly satisfying because we never see bits of electrons; they are either whole or not.

- In 1926, German physicist Max Born took this idea one step further and suggested that we think of the wave of electrons more like a cloud of probability: In spots where the wave's magnitude is large, there is a high probability that the electron will be found there. Where the magnitude is small, we're unlikely to see it, but it's still possible.

- Quantum mechanics, as it stands today, suggests that we can never know exactly where the electron is at a given moment in time *and* how fast it is moving. This paradox was captured by German physicist Werner Heisenberg in his famous uncertainty principle: We can't ever know definitely where an electron is and where it's going, or, more precisely, how fast it's going. The more precisely we define its position, the less well we can measure its velocity, and vice versa.

- Physicist Richard Feynman demonstrated that Heisenberg was right: We can never know both where an electron is and how fast it's going—and, therefore, we can never predict where it's going to be next.

- Feynman demonstrated this simply by trying to locate the electron. One way to find a single electron is to shine a light on it—to bounce photons off of it and measure the trajectory of the photons, then extrapolate back to where the electron must have been.

- When we try to find the electrons as they pass through the slits of the double-slit experiment, by shooting photons at the electrons as they pass through the slits, we find that we end up altering their trajectory and changing the outcome of the experiment.

- No matter how gently we probe the electron to find out its location, we still affect its movement. And we know that the movement is affected because the interference pattern that we expect to see on the photographic plate changes. This interference pattern is in effect a measure of momentum.

- In terms of the actual structure of atoms, or the elements of matter, the Heisenberg uncertainty principle and Feynman's experiment tell us that electrons are bound to the nucleus by the electromagnetic force—they are attracted to its positive core. And now we know that their placement around the nucleus is best described by a cloud of probabilities, rather than by any specific orbital route. Particle waves are waves of probability.

**Applications of Wave-Particle Duality: Lasers**

- We still don't really understand the duality of light, but despite its mysterious nature, we have figured out ways to use it to our advantage. Arguably the most widely used application is the invention of the light amplification by stimulated emission of radiation (LASER).

- When an electron moves from a high-energy state to a lower one, one orbital shell to another, a photon is emitted. So, light essentially is the movement of electrons between different energy states.

- Stimulated emission happens when an incoming photon stimulates the electron to change energy levels and emit another photon. Then, the new photon joins the incoming one, exactly in the same phase. That is, their wavelengths line up; they're in sync. We say that the two photons are coherent.

- This coherence is a fundamental property of lasers and underlies their power: Because the photons are traveling in the same phase, we can control where they can go by making very small slits or openings and restricting their movement. And because all of the light in a laser beam can be restricted to a small area, the intensity, or power per unit area, can be very high.

- Sometimes, an electron jumps from one shell to another—from a lower-energy state to a higher one. When it's at a lower level, it's more stable, closer to the nucleus. And because the electron is closer to the nucleus, the binding force holding it close is greater. These rogue electrons eventually have to move back down to their original lower states, and when they do that, they emit a photon of light.

- Not only do excited electrons emit photons on their way home, but they also encourage other nearby electrons to do the same. An incoming photon stimulates the emission of another photon with the same wavelength and direction. The architecture of lasers takes advantage of these properties.

## Suggested Reading

Feynman, *Six Easy Pieces*.

Schwarcz, *Radar, Hula Hoops and Playful Pigs*.

## Questions to Consider

**1.** How does the uncertainty inherent in the very smallest particles still result in a relatively predictable physical world?

**2.** If particles can be in two places at the same time, then is time travel possible?

# Quanta, Uncertainty, and a Cat

## Lecture 21

You have already learned some of the things we know about quantum mechanics in previous lectures. In this lecture, you will explore what we still don't know, in addition to learning more about what we think we know. In science, it's just as important to know the boundary of our knowledge as it is to know the information itself. Nanoscience often sounds like science fiction, yet many of the most outlandish predictions of quantum mechanics have withstood the tests of time and experimentation.

**What Is Quantum Mechanics?**

- Quantum mechanics is the branch of physics that attempts to explain the behavior of very, very, very small things. The term "quantum" describes a feature of the nanoscopic world that is both fundamental and intuitively puzzling: the fact that some physical properties can only change in discrete amounts.

- For events in our day-to-day world and all the way down to the scale of molecules, classical physics predicts behavior very accurately. But that doesn't necessarily mean that quantum laws aren't still in effect: We just don't see them.

- We experience the macroscopic world as smooth; we can get a little hotter or run a little faster or feel a little sadder. Things seem to be continuous. But that's because the quantum effects occur at the tiniest of scales: We observe light dimming as the Sun goes down, but every photon that hits our retina has a discrete energy associated with it. Our universe is granular, not smooth. But we operate on a scale at which that granularity appears to be smooth and the laws of classical physics apply.

- Quantum theory was born because these laws don't seem to apply to the subatomic world. Classical physics made predictions that repeatedly were proven wrong during experimentation. The

mathematical formulas and rules derived from the principles of quantum mechanics, in contrast, have made some of the most precise and accurate numerical predictions in the history of science. The soul of quantum mechanics is math.

**The Copenhagen Interpretation**

- Despite the lack of consensus among physicists, the Copenhagen interpretation is both one of the earliest and still one of the most popular ways to think about quantum mechanics. The tenets were proposed by Niels Bohr, Werner Heisenberg, and others—not by Albert Einstein, however.

- In simple terms, the interpretation suggests that quantum mechanics is not built to describe objective reality, but that the act of measuring quanta forces the set of probabilities to randomly assume only one possible value. In a nutshell, by observing the quantum world, we force it to choose one of several possible paths.

- There are several principles that have been attributed to the Copenhagen interpretation. First, the best way to describe a system, or a set of interacting components, is by using a wave function that represents the state of the system and changes with time—except when a measurement is made. A measurement causes the waveform to collapse instantaneously into an observable, measureable state.

- Second, the description of nature is probabilistic at its core—we can't specify every aspect of it at once, just as we can't simultaneously know the position and velocity of an electron. So, we use probabilities; we describe elements of the system in terms of their likelihood.

- Third, we cannot know all of the values of the properties of a system at the same time. This is the Heisenberg uncertainty principle, restated a different way. Those values that we don't know precisely are described in terms of probabilities.

- Fourth, matter acts both like a wave and a particle, demonstrating wave-particle duality. When we experiment on matter, we can observe particle-like or wavelike behavior and, sometimes, both at the same time.

- Fifth, when we measure things, we use the instruments of classical physics, and therefore we only measure classical properties like position and momentum.

- Finally, in large systems, the quantum mechanical description will converge upon the classical physics description—this is what's called the correspondence principle. What classical physics predicts becomes the same as what quantum mechanics predicts, when the components of the system are large.

**Schrödinger's Cat**

- There are ways in which the Copenhagen interpretation helpfully distills the core paradoxes of quantum mechanics. The most famous one involves a cat—Schrödinger's cat in particular.

- Erwin Schrödinger, an Austrian physicist, wanted to demonstrate one of the paradoxes that the Copenhagen interpretation presents—that is, that this interpretation seems to contradict common sense. Namely, he wanted to explore the idea that if a waveform describes a system, and it collapses into a point when we measure it, then it must be measured at all possible points at the same time. For example, if we predict the velocity and position of an electron using a probability, this implies that it is in all possible places at once.

- To illustrate this paradox, he decided to use a cat in a thought experiment. Schrödinger put a theoretical cat into a theoretical box, along with some radioactive material and a Geiger counter (a tool for detecting radioactivity). If the Geiger counter detects radioactive decay, then it will trigger the action of a hammer, which will break a flask containing a poison, which would kill the cat.

- If you leave the cat in the box indefinitely, it will certainly die because the radioactive material will certainly have time to decay. But if you only leave the cat in the box for an hour, it may or may not live through the experience. It's possible that the radioactive material will decay. But it's also possible that it won't.

- When you seal the box, you don't know if the cat is alive or dead—it could be killed at any moment. But we can't see it, so we don't know. It's a bit like a particle, which can exist in multiple states. But because we force it to choose a state when we observe it, we can't observe it without affecting the outcome. So, what state is it in? Is the cat alive or dead?

- According to the Copenhagen interpretation, in the box, the cat is both alive and dead at the same time. The particle is in all possible states at once. When we observe it, we force it to take one of the possible states.

- The state of existing in all possible states at the same time is called superposition. All of the possible states are described by the object's wave function. When we make an observation, superposition collapses, and the object is forced into one of the states along its wave function.

- Superposition is an important corollary of the quantum theory in general and the Copenhagen interpretation in particular. The idea behind superposition is that if an object, particle, or system can exist in multiple states—arrangements of particles or fields—then the most general state, or the default state, is a combination of all of these possibilities.

- Some states might be more likely than others, due to inherent stability or other factors, so the amount that each state contributes can be thought of as a weighted value: More likely states are given more weight.

- The double-slit experiment, in which the wave-particle duality of light was observed, is also “proof” of the principle of superposition. Until we actually try to observe the electrons directly, they pass through both slits at the same time.

- A related concept from quantum mechanics that comes into play when we talk about superposition is entanglement, which is the capacity of paired particles to be linked even if they are separated in space.

- Quantum entanglement is a direct result of superposition, because before we observe the particles, they are described by a wave function that incorporates all of their possible states. But when we measure one of them, the outcome of the other is instantly guaranteed.

- If you’re uncomfortable with the idea that two things can be linked to each other even when miles apart, you’re not alone. Part of the problem with it is that it seems to imply that information can travel faster than the speed of light—because, otherwise, how could one particle determine the state of the other, instantly, from miles away? However, not only does entanglement make accurate predictions that concur with observations, but it also already has some applications.

**Applications of Quantum Mechanics: Nanoscience**

- When you consider how truly small the nanoscopic world is, it’s hard to believe that we’ve learned to build and manipulate things of that size. Some nanotechnology is one five thousandth the size of a sperm cell, one of the smallest cells in the human body.

- At that level of tininess, substances sometimes behave just as erratically as quantum mechanics predicts. But we can do some things there that are simply impossible on a macro scale.

- For example, there is currently no way to teleport humans through physical barriers, but we can send one electron through—using

© Stacey Newman/iStock/Thinkstock.

**In classical physics, a ball rolling up a hill can only roll upward until it runs out of energy, at which point it will roll back down the hill.**

quantum tunneling, which is the process by which an electron passes through a barrier that classical physics predicts it shouldn't.

- To understand this conundrum, consider what classical physics would say about a ball trying to roll over a hill. Without sufficient energy, or a push, the ball wouldn't be able to do it; it might roll up part of the hill, but as soon as it runs out of energy, it would roll back down. If the hill were a wall instead, it would simply bounce off the wall, not through it.

- But according to quantum mechanics, a quantum particle shows wave-particle duality, and its position and momentum can't be precisely determined—as the Heisenberg principle tells us. So, we can't say with 100% certainty that the ball, if it were a quantum particle, can never pass through the wall or over the hill. There's a very small probability that it could tunnel through the hill or wall by borrowing energy from its surroundings, giving it back on the other side.

- Quantum tunneling is thought to perhaps underlie radioactive decay, among other applications. Spontaneous DNA mutations also might result from quantum tunneling, but using protons rather than electrons. There is even a new microscope called the scanning tunneling microscope that relies on quantum tunneling through a metal surface.

- What might the future of nanotechnology hold? What if we could build nanorobots that could perform delicate surgical procedures in our bodies without the need for a scalpel? Imagine if we could just drink a solution that contains the robots and let them do their work. Surgical scars would be a relic of history.

## Suggested Reading


Gilder, *The Age of Entanglement*.

Greene, *The Elegant Universe*.


## Questions to Consider

**1.** Why is using quantum mechanics to explain things we don't yet understand fully, such as human consciousness, so compelling?

**2.** What are the implications of superposition? How can Schrödinger's cat be both dead and alive at the same time?

# String Theory, Membranes, and the Multiverse
## Lecture 22

Theoretical physics is arguably the most imaginative of all the sciences, and for many people, its thought experiments, its mind-bending theories, and their implications can be unsettling. Still, it's worthwhile to delve into these ideas even in a fairly shallow sense, because, in the true spirit of science, they raise far more questions than they answer. This lecture will introduce you to string theory and a few theories that are alternatives to the standard model of physics.

**Essentials of String Theory**

- Despite the fact that the standard model of physics, with its quarks and bosons, is an accurate description of the subatomic space, there are still a few things that it fails to explain, including gravity. Is the very premise that the smallest things are point-like particles, so small that they virtually have no volume or shape, incorrect?

- What if particles were instead made up of tiny vibrating strings? You could have open strings or closed strings that form a loop, and the different ways in which they vibrate could imbue all the particles we know of with different properties.

- Outlandish as it might be, string theory and other alternatives to the standard model provide us with provocative implications and stimulate creative thinking. Even if ultimately each of these theories is shown to be inaccurate, understanding why these theories have gained ground in the past few decades will tell us that much more about the world we inhabit.

- String theory is fundamentally different from the standard model because it provides a putative explanation for *why* the building blocks of subatomic particles might be shaped like strings. The standard model does no such thing; it just predicts that these

© Andrew Gentry/iStock/Thinkstock.

**The vibrational patterns of violin strings can be used as an analogy to explain string theory.**

particles must exist, and because we've observed them, it treats the subject as resolved.

- String theory provides an explanatory foundation that is either right or wrong, but it's not infinitely modifiable. Strings have certain properties that make them good candidates for the basic units of all matter, unlike any other putative shape.

- Let's take violin strings as a starting point. A violin string is tied to the instrument at two points. But in between those two points, there is nearly an infinite number of resonances that the string is capable of creating, simply by changing its vibrational patterns. A violinist can play many different pitches and make many different sounds with only one string.

- So, too, can the strings in string theory: The different patterns of vibration of a fundamental string can lead to different forces

and masses. So, the different subatomic particles are simply manifestations of different vibrational patterns of strings.

- If you pluck a string on a violin lightly, it vibrates just a little. If you pluck it with a lot of force or energy, it vibrates more vigorously: The amplitude of the sound wave that the vibrating string generates is larger, the difference between the peaks and valleys is greater, and the wavelength is shorter—there is less distance between two peaks.

- The same is true of subatomic strings: The shorter the wavelength and the larger the amplitude, the more energy the string has. As Einstein taught us, there is a direct relationship between mass and energy; greater energy equals greater mass. So, the subatomic mass of a particle is determined by how vigorously a string is vibrating. The more frantic the vibration, the more mass the particle has.

- And here's where string theory seems to outdo the standard model: The more mass the string-generated particle has, the greater the gravitational force that it exerts. In fact, all of the fundamental forces, including gravity, can be translated in terms of the vibrational patterns of the strings.

- In the standard model, each subatomic particle is different—made from different "stuff." In string theory, they are all made of strings; some just vibrate one way and some another, giving them different properties. This is why string theory often includes musical analogies—music of the spheres, or the cosmic symphony: Different elementary particles are simply different notes played on the same strings.

- A melody is made up of musical phrases strung together. Each phrase is made up of chords or other clumps of notes. Each note might have a pitch value and a time value—how high it is and how long it lasts—but you can't reduce it more than that without losing the essence of what a note is. In the same way, you can't reduce

strings to anything else. If they were composed of other things, they wouldn't be the fundamental building blocks of matter.

- We all experience the world in 4 dimensions: 3 describing space and 1 describing time. Spatial dimensions are roughly translated into left-right, backward-forward, and up-down, while time moves in one direction, from the past to the future. This 4-dimensional representation of our world and our experiences is so ingrained that it's very difficult for us to imagine the universe described in any other way.

- But string theory doesn't quite work in 4 dimensions. Gravity, when Einstein described it, seemed to be a mysterious force that somehow affected large bodies from very long distances, with only space separating them. Einstein's genius was in suggesting that the space itself was the material on which the force acts.

- That space and time can be thought of as a large, flat sheet, and the bodies that are very heavy create depressions in that sheet that affect the movements of nearby smaller bodies. And according to Einstein, the fabric of space-time is constant; it's a smooth continuum that only very big things, like planets and stars, can warp.

- Einstein had a problem with the fuzzy nature of quantum mechanics because if space-time is the fabric of the universe, and it's smooth and constant, how is it possible that the smallest things, those elementary particles that make up everything, are always moving about and shifting in ways that we can't predict or observe?

- String theory tries to reconcile the two theories by suggesting that the properties of the strings that make up the universe can yield gravity and all of the other forces. The different ways in which the strings vibrate correspond to the particles of the standard model, including the force carriers. For the first time, the 4 forces are unified by the properties of vibrating strings.

- But it also goes a step further: What if, in the fabric of space-time, there were other tiny dimensions that we couldn't see? For example, what if, deep in the nanoscopic world, strings could move in the typical 3 dimensions and the fourth, the time dimension, but also in other dimensions around little loops at the edges of where our familiar 4 dimensions intersect?

- String theorists often suggest that we picture these alternative dimensions like the circumference of a thin wire: If you look at the wire from far away, as the nanoscopic world looks to us, it only has 2 dimensions—it looks like a flat line. But if you zoom in, you can see that it's actually a tube, and things that interact with it, just like an ant on a telephone wire, can move in the 2 dimensions, but it can also crawl around the circumference or around the wire.

- String theory suggests that that's how tiny the new dimensions might be—so small that we can't observe them. And there isn't just one, like the circumference of the wire; there are 6 of them. So, for the computations and predictions of string theory to be correct, there have to be 9 space dimensions: the 3 that we can observe and 6 that we can't, and 1 time dimension for a total of 10.

- These extra dimensions intertwine with each other to create an intricate shape. Just like a cello and a violin make different sounds because the instruments have different shapes, so, too, might the vibrations of strings depend on the intricate shapes of these tiny dimensions.

**M-Theory and Black Holes**

- A few years after string theorists posited that 10 dimensions were needed to make the equations of the theory work out, another version of the theory suggested that 11 dimensions were needed to account for some of the features of gravity.

- Then, another version was proposed that also seemed to be correct, even though some of the equations were tweaked. Then, another and another were proposed until there were 5 versions of string

theory, all with different equations, and all of which seemed equally plausible or correct.

- In 1994, string theorists came upon a potential solution: What if all the theories were actually correct and simply described the world from different points of view? The 5 theories might be the same thing, from different perspectives.

- So, they proposed a unifying theory that incorporated all 5 and called it M-theory, which is generally thought of as suggesting that the strings are slices of a 2-dimensional membrane that vibrates in an 11-dimensional space.

- At the moment, M-theory is still more of an idea than a theory, even though many of its details have been outlined mathematically. Some leading physicists claim that it will remain ultimately untestable, but for some people, even the thought of testing the theory brings on fears of world-ending black holes created by particle accelerators.

- You can think of black holes as a thread that pulls together the laws of classical physics, including Einstein's theories of relativity and quantum mechanics: Large black holes have properties that are similar to elementary particles, but they obey the laws of classical physics, while micro black holes, the kind that are formed in the dimensions that we can't see according to string theory, obey the laws of quantum mechanics. By understanding black holes, we learn about elementary particles, and by learning about elementary particles, we might begin to understand black holes.

**Alternative Theories and Their Implications**

- At the very center of a black hole is what physicists call a gravitational singularity, which is a curve in space-time that reaches infinity. Einstein's theory of general relativity predicts a gravitational singularity just before the big bang, which keeps the big bang shrouded in mystery. By extending the curvature of space-time to infinity, a gravitational singularity remains unexplained. String theory and M-theory have the same problem.

- But an alternative theory, proposed in the 1980s, provides an attractive answer: What if the core of the big bang is not nothing, but a very tiny, hugely compressed universe? And what if the big bang isn't really a bang, but more like a big bounce? In other words, the universe gets very, very small, and then something happens to cause it to expand once again. This idea is a corollary of yet another theory, called loop quantum gravity.

- Loop quantum gravity opens the door to the idea that there are—or were, or will be—other universes besides ours. But unlike in M-theory, there is no need for extra dimensions; the 4 that we can observe work just fine.

- Einstein's problem with Schrödinger's cat was that uncertainty is not a solution. So, the many worlds interpretation suggests an alternative: What if, during each instance of uncertainty, both consequences happened, but in different universes, so the cat was alive in one and simultaneously dead in the other? The many worlds interpretation has a substantial following in theoretical physics, and its implications continue to occupy some of the greatest minds of our time.

## Suggested Reading

Greene, *The Elegant Universe*.

Kaku, *Physics of the Future*.

## Questions to Consider

1. If the universe isn't made up of tiny vibrating strings, what other shapes or objects could be the smallest bits of matter?

2. Are our lives less meaningful if there are parallel universes? Or would evidence for multiverses be evidence for eternal life?

# Emergence—Simple Rules, Complex Systems

## Lecture 23

The notion that the whole is greater than the sum of its parts is at the core of a relatively new field of science dedicated to studying complex systems called emergence. By observing the behavior of slime mold, ants, birds, fish, and other animals, we can begin to see how order can emerge out of chaos and how the collective can be smarter or more effective than the individual. As you will learn in this lecture, the key to understanding emergence is to uncover the rules that the individuals follow.

**General Principles of Emergence**

- Slime mold has no central executive: It's made up entirely of individual single-celled organisms; it's a moving clump of individuals, not an organized team of parts. But when food is scarce, the individuals gather and form an aggregate that behaves as a single organism. When food is more abundant, the "it" separates into a set of "theys": The aggregate breaks down, and the individual organisms go their separate ways. As such, slime mold is a simple model for the coordination of group behavior.

- After decades of research, which started with the pacemaker idea that there is a hierarchy to the slime mold, Evelyn Keller and Lee Segel suggested that all the individual single-celled organisms in the mold might be secreting a pheromone, or a type of hormone that triggers a social signal in members of the same species, in response to environmental conditions.

- When conditions are right, they hypothesized, the individuals secrete more of the pheromone. Then, they follow the trails of the pheromone that they come across from other individuals. Eventually, they begin to find each other, clustering in small groups at first and amplifying the pheromone signal. The small groups become medium-sized groups, and more cells are attracted as the amount of pheromone increases.

- At some point, the community turns into a large slime mold conglomeration, all without any need for a hierarchy or command center. This is a classic case study of bottom-up organization, as opposed to top-down, as is the case in hierarchies.

- Once scientists recognized this phenomenon, they began to find it in many places, including in the formation of city neighborhoods, ant colonies, swarming behavior of birds and fish, and even the human brain.

- What is the common element in these complex systems? The building blocks are relatively simple, but they mass together to overcome an obstacle or function as an effective unit, without the leadership of a central executive. They display the characteristics of emergence: Low-level rules lead to high-level sophistication, or the rise of complex systems out of relatively simple interactions.

- When we want to understand a system as complex as the brain or the stock market, we can start with a simple system to extract some general principles. This area of inquiry has become known as the field of emergence.

- In emergence, simple agents together perform complex feats that are impossible for any of the individuals to accomplish on their own. But it's not just the sheer number of simple building blocks that is the key: Without more, the building blocks' behavior is more likely to be chaotic rather than organized, competitive or counterproductive rather than harmonious. Rather, the interaction between a set of simple rules that govern the behavior of the agents and their large numbers yields a whole greater than the sum of its parts.

- Emergence can come in one of two forms: emergent behaviors or emergent structures. Behaviors describe the actions of a group of individuals, such as the swarming of schools of fish, or birds, or the work of an ant colony. Emergent structures are the actual patterns observed: the layout of a city neighborhood, or the shape of the school of fish, or the ant colony itself.

- Emergence has certain properties that we're just beginning to understand. By grasping its sometimes counterintuitive principles, we can begin to predict the behavior of complex systems, even systems like the stock market, which seem to defy much of what we expect.

- For example, emergence appears to reverse entropy. Emergent behavior is sometimes described as a spontaneous move toward order. The idea that single-celled organisms like the constituent parts of slime mold come together and behave as an organized unit seems to defy the second law of thermodynamics, which states that systems move toward a state of disorder or chaos rather than order.

- The second law states that the total order in the system must decrease—but not where and how. For example, when food is scarce, the slime mold collects individuals, decreasing its entropy. But then, the aggregate is more efficient at extracting energy from the forest floor, thereby breaking down molecules and increasing entropy in the ecosystem at large.

- There are several other properties or rules that emergent behaviors or emergent structures seem to follow. For example, more is different: A few individuals don't seem to show the emergent pattern, but thousands do. It's only by observing the system as a whole that the behavior or structure comes to light.

- Also, ignorance is useful: The individual components should be fairly simple and/or stupid. That's why we build computers using the binary system of ones and zeros. If the components are too smart, they might begin to make their own decisions, throwing the system out of whack. However, there's some question as to whether this rule always applies.

- Emergence also depends on, or encourages, random encounters: Without haphazard interactions, the system might not be able to find new solutions to a problem. A slime mold might not find a path that leads to food. An ant colony might not notice an invader.

- Another property is that there are patterns in the signs that an emergent system's constituent parts exchange with one another. The pheromone trail of a slime mold has a gradient that the aggregate follows—more pheromone means more food. Ants communicate with each other using a limited number of signals.

- Finally, local information can lead to global wisdom. Slime mold individuals talk to their neighbors, not some central executive, but the result is behavior that helps all of them.

**Ant Colonies**

- We like to think of humanity as the most successful species on Earth, but there's an argument to be made for the superiority of insects. We might be able to speak, read, and write, but honeybees can dance their way out of a sticky situation, and ants have a language of chemicals that includes signals for food sources, task assignments, and even a signal indicating that it's time to bury a



**Emergence principles can be applied to the behaviors of ant colonies.**

friend. Arguably, insects of the hymenoptera order—including ants, bees, and wasps—rival the social complexity of humanity.

- It's easy to kill an ant. On its own, the little creature is simple and stupid. But ants live in colonies, some of which contain upward of 20 million individuals, and they're not defenseless when they unite. Ant colonies are arguably better organized than many of our cities, with strategically placed cemeteries, trash heaps, and even an underground bunker to protect their queen.

- The queen moniker is deceptive: Unlike in a human monarchy, she wields little power over her subjects. She can't tell them what to do, or organize their activity, or even influence their behavior aside from giving birth to new offspring. Instead, she's entirely at the mercy of the colony.

- What's really amazing about ants is just how complex and organized a colony can be without a brain or a central leader. And by observing the movements and behaviors of the individuals in a colony, we can begin to understand how the principles of emergence can help the lowly ant achieve great things.

- The emergence principle that local information can lead to global wisdom applies to ants. Ants think and act locally, but the conglomerate of their actions improves the conditions for the entire colony.

- Harvester ants, for example, need to bring in enough food for the entire colony, but the size of the colony changes with time. They also need to make sure that there is just enough (but not too much) food stored in the nest, that their surrounding area has enough food to bring into the nest, and that there aren't any other ant colonies in the vicinity. And they need to dispose of the colony's dead bodies at the maximum distance from any colony entrance.

- No individual ant is capable of making these assessments and adjusting the colony in response, but every ant speaks the language

of pheromones: 10 or 20 chemical signals that can tell an ant who is on foraging duty, where the food is, whether there is danger, and when it's time to bury a fallen comrade. These chemical signals are on their bodies, as well as in the trails that they leave behind as they go about their business.

**Swarm Behavior**

- It's not just ant colonies that can teach us about emergent behaviors. One of the most amazing wildlife experiences is to observe swarm behavior: swarms of insects, schools of fish, flocks of birds, or herds of mammals all moving in one complex, coordinated group.

- Like slime mold and ant colonies, there is no central organizing force. Rather, the group is made up of relatively simple individuals, all of approximately the same size, who follow a set of rules. From this set of rules emerges a coordination that is not possible from any single individual.

- Applied mathematics has made a huge leap in our understanding of flocking behavior. In 1986, a computer programmer named Craig Reynolds created an artificial simulation of birds flocking.

- To simulate flocking, he programmed up simple birdlike objects and instructed them to behave according to three rules: steer clear of local flock-mates, thereby avoiding crowding; steer toward the average direction in which local flock-mates were heading; and steer toward the average position or the center of mass of local flock-mates. He called these rules separation, alignment, and cohesion, respectively.

- This very simple framework proved to be remarkably accurate in simulating the flight of a flock of birds. And by extending it to add obstacle avoidance and goals, such as finding food or avoiding predators, the program is nearly perfect.

- Just like birds, fish that swarm in schools follow simple rules for navigating as a group. About a quarter of all fish stay together,

or shoal, for their entire lives. Half of all fish shoal for part of their lives.

- Arguably the most established explanation for herd behavior is the selfish herd theory, which suggests that herd behavior is largely selfish rather than social. The idea is that individuals within a population seek to reduce the risk of being eaten by predators by placing tasty alternatives between themselves and their hunters. Because predators generally attack the closest prey, ensuring that there is a member of your species between you and the predator is a fairly effective way of protecting yourself.

- How do shoaling, flocking, or other herd behavior that involves complicated coordination emerge? Taking fish as the example, many young fish have to practice schooling in small groups before they can join a large group. This practicing behavior is instinctive, rather than being taught by their elders.

## Suggested Reading

Gordon, *Ant Encounters*.

Johnson, *Emergence*.

## Questions to Consider

**1.** What are some other complex systems whose rules you'd like to understand that affect your life?

**2.** What would the consequences be if we could predict the stock market?

# Order out of Chaos

## Lecture 24

Many of the complex questions that emergence poses and seeks to answer have no sharply defined boundaries: When do neurons become a brain? How many residents make a neighborhood? The same blurry edges characterize many of the questions that have been explored in this course: What is life? Where is the line between the macro and nanoscopic worlds? Have we found the smallest particles yet? Science can't yet provide perfect definitions for many of the big topics that fascinate us, but we can set the boundaries for the purpose of an experiment and discover an answer to a specific question.

**Artificial Intelligence and Social Robotics**

- In the 1960s, as computer science was in its infancy, John Holland was obsessed with the problem of how complex behavior can emerge from simple rules. Having just obtained the first Ph.D. in computer science from the University of Michigan, he set out to develop a computer program that was capable of learning.

- At the time, the best-known scientific theory of how order can emerge from chaos was evolution, specifically the mechanism of natural selection. So, Holland borrowed from the ideas of Darwin to pioneer a type of computer program called a genetic algorithm (GA), which is a type of shortcut, or heuristic, that uses the principles of natural selection to optimize searches.

- You can think about learning as essentially a search process: There's a problem, and you need to find a solution. Once you do, you've learned something—that your problem can be solved with the solution that you just found.

- For example, if you want to learn how to navigate through a maze, you first try one direction and then another until eventually you find

a path that takes you to where you need to be. You learn to navigate the maze so that next time, you can just take that one path.

- Genetic algorithms work in a similar way, except that natural selection is applied by specifying the population of potential solutions—all the different ways of going through the maze, for example, and then, over many trials or iterations, the best solution "wins," or outlasts all the suboptimal ones.

- Every genetic algorithm has at least two components: a representation of the environment in which the solution will be found and a fitness function to evaluate the solutions. Each potential solution has a particular set of properties, akin to its genetic code, that can undergo mutations.

- And the algorithm works by creating multiple generations of these solutions, whose fitness or solution strength is then evaluated. The algorithm stops when either a preset number of generations has been created or a solution that reaches a target fitness level is found.

- In successful algorithms, the solutions that are generated in the later generations are qualitatively different from the potential solutions suggested to begin with: The information evolves. Something better comes from simple inputs; the computer learns.

- In the 1980s, shortly after Holland showed the world how useful GAs might be, computer scientists David Jefferson and Chuck Taylor at UCLA designed a program that utilized the principles of GAs to simulate how ants might forage for food. They found that the line between real- and virtual-world learning began to blur when adaptations that were not programmed into the software emerged from the rules that were set out—virtual ants learned and evolved.

**The Emergence of Theory of Mind**

- Nowadays, genetic algorithms and other computer programs can perform many feats that our own brains are simply incapable of.

But there are still a few things that we're pretty good at, even at a young age, that machines have found too taxing.

- One of these things is sentience: understanding, or believing, that we are conscious. For most of us, consciousness is still the most important thing that separates us from artificial life. A computer, no matter how well it might be able to solve problems, still can't feel pain. But what makes *us* self-aware, when computers, which can outperform us in so many ways, aren't?

- Each one of our bodies is made up of fairly simple cells, all acting with simple rules to follow, many of which are written into our genetic code. None of the individual cells can think or make high-level decisions. And out of that chaotic cacophony, our complex natures emerge. We're capable of so much more than the sum of our parts—or so it seems, at least.

- Despite the fact that none of our neurons can think on their own, we experience our minds as integrated, and we have a bizarre self-awareness: We can introspect or observe at least some small part of our thinking. It turns out that we're far less aware than we think of the computations that our brains make that lead to our decisions and, ultimately, our behavior.

- How might consciousness have emerged from this tangled mass of stupid cells? Our brain cells have a binary nature: They can either fire an action potential and propagate a signal, or not. Neurotransmitters are the pheromones of neurons: They can influence the behavior of individual cells, and they have information both in terms of their presence or absence and in terms of their gradient—how much of each type are floating around synapses at any given time.

- Just like ants, different neurons have different functions: Some translate information from the senses into electrical signals, as in the eyes and inner ear; some pass information along from peripheral body parts; some inhibit firing in the hippocampus to prevent

seizures; and some tick like clocks, keeping your body aware of the time of day.

- Many of these neurons have jobs assigned from birth, although they can make some changes as their circumstances change with age, injury, or training. And some neurons are specifically involved in helping you understand the intentions and beliefs of people and animals around you.

- These latter neurons were first discovered by a group of Italian neuroscientists headed by Giacomo Rizzolatti, who were interested in the brain cells that are involved in motor actions like grasping a cup or eating a banana. So, they started recording firing patterns, or electrical signals from neurons, in monkeys while the monkeys were performing these tasks. They found cells that would fire when the monkeys were engaged in a specific task.

- But they also found that those same cells sometimes fired when the monkey was watching someone else perform that task. They called these cells "mirror neurons" because they seemed to be mirroring what they were seeing.

- There's nothing special about these neurons in terms of their shape or size or any other distinguishing features; the only thing that makes them mirror neurons is what they seem to respond to—what stimulates them. And these mirror neurons make up about 10% of the neurons in the parts of the monkey's brain where they are normally found.

- Mirror neurons have also been found in humans. They seem to perhaps play a role in how we understand the intentions of others—how we develop a "theory of mind," or the idea that other people have thoughts, desires, and goals, just like we do.

- Even if these particular cells ultimately aren't the neural basis of theory of mind, they provide a good case study for how sentience and self-awareness might emerge from simple cells.

- No single mirror neuron is actually conscious; no single cell "knows" that another being has intentions. Instead, the information that these cells carry can help the colony, or the brain, extract another person's intentions. It's an emergent property.

- How might this work? Individual mirror neurons might provide the information that a monkey or baby or any one of us needs to figure out how to mimic or mirror an action that we see someone else perform. When we mimic that action, we put ourselves in the other person's shoes—and we begin to discover what it might be like to be that person.

- These neurons, and the system that reads their information, seem to be involved in the development of theory of mind. We don't know exactly how this ability to consider the minds of others came about, but some clues can be found in the ways in which we've evolved to live in social groups.

- Social interactions can extend your life span, just as the lack of social support can hasten death. The vast majority of humans live in extended groups, most of us with complex social systems. And this trait makes us unique among apes: Only chimpanzees live in mixed-sex social groups like we do.

- Some psychologists go so far as to suggest that the exponential increase in the ratio of brain to body size that characterizes our species emerged as a result of our living in groups. The social brain hypothesis suggests that we evolved into our intelligent selves not in response to ecological problems, but so that we can all get along.

- The idea is that theory of mind was selected for as we began to live in groups. If that's true, then those neurons that aid us in forming our own mental picture of the thoughts and intentions of others might represent the building blocks of self-awareness. If we learn that others have a mind, can we not turn those same cells inward and observe our own minds?

© Jupiterimages/Stockbyte/Thinkstock.

**People who engage in social interactions live longer, on average, than people who do not.**

- Social complexity possibly led to the explosion in brain size by simply increasing the number of computations that we need to make in order to understand multiple other minds. If there's a module in the brain that is made up of cells like mirror neurons, or some similar equivalent, and it works fairly well at analyzing one person's mind, maybe understanding many people's minds simply means adding more modules.

- At some point, however, we reached a limit: It turns out that most of us have a social circle that caps out at about 150 people. This is called Dunbar's number, after Robin Dunbar, who first proposed the social brain hypothesis. The larger the social group, the bigger the neocortex, and that includes mirror neurons.

- The number 150 seems to represent the upper limit of people with whom we can maintain stable relationships, not the total number of people that we know. Some other anthropologists have come up with larger numbers, but the idea that there is a limit is fairly well accepted.

- But as a civilization, we've managed to overcome this limitation of our brains: not by continuing to increase neocortex size, but by linking up brains into larger networks. In our day, those networks have begun to expand dramatically.

**From Brains to Cities**

- There came a time, likely when we figured out how to domesticate animals and grow plants, when we started to gather in groups that were larger than 150 people. We started to live in cities, with tens of thousands of individual minds to keep track of. And our poor neurons simply couldn't keep up.

- So, a new level of organization had to emerge: Instead of neurons being the building blocks, or the ants in the colony, we, ourselves, became the simplest units. We started living in neighborhoods, and now our colony size exploded.

- Think about how this is happening in your own life. No matter where you live, your neighborhood, your town, your city has a personality—traits that separate it from other places to live have emerged as distinct.

## Suggested Reading

Kurzweil, *How to Create a Mind*.

Mayer-Schonberger and Cukier, *Big Data*.

## Questions to Consider

1. If consciousness can emerge from the activity of many millions of simple neurons, what emergent properties do our other organs show?

2. What are some of the emergent systems that surround you? How has your neighborhood changed in response to the activity of the humans that inhabit it?

# Bibliography

Anderson, John. *Fundamentals of Aerodynamics*. New York: McGraw-Hill, 2010.

Bergstrom, Carl T., and Lee Alan Dugatkin. *Evolution*. New York: W. W. Norton & Company, 2011.

Carlson, W. Bernard. *Tesla: Inventor of the Electrical Age*. Princeton, NJ: Princeton University Press, 2013.

Carnot, Sadi. *Reflections on the Motive Power of Fire and Other Papers on the Second Law of Thermodynamics*. Translated in 2005. Mineola, NY: Dover Publications Inc., 1890.

Cengel, Yunus, and John Cimbala. *Fluid Mechanics*. New York: McGraw-Hill, 2009.

Cengel, Yunus, and Michael Boles. *Thermodynamics: An Engineering Approach*. 7th ed. New York: McGraw-Hill, 2010.

Cox, Brian, and Jeff Forshaw. *Why Does* $E = mc^2$ *and Why Should We Care?* Philadelphia: Da Capo Press, 2010.

Darwin, Charles. *On the Origin of Species: By Means of Natural Selection*. New York: Appleton, 1860.

Doidge, Norman. *The Brain That Changes Itself*. New York: Penguin, 2007.

Drummond, James E. *Plasma Physics*. New York: McGraw-Hill, 1961.

Feynman, Richard P. *Six Easy Pieces: The Essentials of Physics Explained by Its Most Brilliant Teacher*. New York: Basic Books, 1963.

Field, Simon Quellen. *Culinary Reactions: The Everyday Chemistry of Cooking*. Chicago: Chicago Review Press, 2011.

Fleisch, Daniel. *A Student's Guide to Maxwell's Equations*. New York: Cambridge University Press, 2008.

Foer, Joshua. *Moonwalking with Einstein: The Art and Science of Remembering Everything*. New York: Penguin, 2011.

Gilder, Louise. *The Age of Entanglement: When Quantum Physics Was Reborn*. New York: Vintage Books, 2009.

Gordon, Deborah. *Ant Encounters: Interaction Networks and Colony Behavior*. Princeton, NY: Princeton University Press, 2010.

Gray, Theodore, and Nick Mann. *The Elements: A Visual Exploration of Every Known Atom in the Universe*. New York: Black Dog & Leventhal Publishers, 2012.

Greene, Brian. *The Elegant Universe: Superstrings, Hidden Dimensions and the Quest for the Ultimate Theory*. New York: W. W. Norton & Company, 2000.

Hawking, Stephen. *A Brief History of Time*. New York: Bantam Books, 1998.

Horowitz, Seth S. *The Universal Sense: How Hearing Shapes the Mind*. New York: Bloomsbury, 2013.

Johnson, George. *The Ten Most Beautiful Experiments*. New York: Knopf, 2008.

Johnson, Steven. *Emergence: The Connected Lives of Ants, Brains, Cities and Software*. New York: Scribner, 2002.

Kahneman, Daniel. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.

Kaku, Michio. *Physics of the Future: How Science Will Shape Human Destiny and Our Daily Lives by the Year 2100*. New York: Doubleday, 2012.

Kamkwanba, William, and Brian Mealer. *The Boy Who Harnessed the Wind: Creating Currents of Electricity and Hope*. New York: Harper Perennial, 2010.

Kean, Sam. *The Disappearing Spoon: And Other True Tales of Madness, Love and the History of the World from the Periodic Table of the Elements*. New York: Little Brown & Company, 2010.

Kurzweil, Ray. *How to Create a Mind: The Secret of Human Thought Revealed*. New York: Viking, 2013.

Langeweische, Wolfgang. *Stick and Rudder: An Explanation of the Art of Flying*. New York: McGraw-Hill, 1990.

Livingston, James D. *Driving Force: The Natural Magic of Magnets*. Cambridge, MA: President and Fellows of Harvard College, 1996.

———. *Rising Force: The Magic of Magnetic Levitation*. Cambridge, MA: President and Fellows of Harvard College, 2011.

Livingstone, Margaret. *Vision and Art: The Biology of Seeing*. New York: Harry A. Abrams, 2002.

Macknik, Stephen, Susanna Martinez-Conde, and Sandra Blakeslee. *Sleights of Mind: What the Neuroscience of Magic Reveals about Our Everyday Deceptions*. New York: Henry Holt & Company, 2011.

May, Brian, Patrick Moore, and Chris Lintott. *Bang! The Complete History of the Universe*. London: Carlton Books, 2012.

Mayer-Schonberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. New York: Houghton Mifflin Harcourt Publishing Company, 2013.

McGee, Glenn. *Bioethics for Beginners: 60 Cases and Cautions from the Moral Frontier of Healthcare*. Oxford: Wiley & Sons, Inc., 2012.

McKenna, Maryn. *Superbug: The Fatal Menace of MRSA*. New York: Free Press, 2011.

Merzenich, Michael. *Soft-Wired: How the New Science of Brain Plasticity Can Change Your Life*. San Francisco: Parnassus Publishing, 2013.

Ouellette, Jennifer. *The Calculus Diaries: How Math Can Help You Lose Weight, Win in Vegas and Survive a Zombie Apocalypse*. New York: Penguin, 2010.

Patrick, Sean. *Nikola Tesla: Imagination and the Man That Invented the 20th Century*. Clearwater, FL: Oculus Publishers, 2013.

Pollan, Michael. *The Botany of Desire: A Plant's Eye View of the World*. New York: Random House, 2002.

Reece, Jane, Lisa Urry, Michael Cain, Steven Wasserman, Peter Minorsky, and Robert Jackson. *Campbell Biology*. 9th ed. San Francisco: Benjamin-Cummings Publishing Company, 2010.

Sagan, Carl. *Broca's Brain: Reflections on the Romance of Science*. New York: Random House, 1980.

Schwarcz, Joe. *Radar, Hula Hoops and Playful Pigs: 67 Digestible Commentaries on the Fascinating Chemistry of Everyday Life*. New York: Henry Holt and Company, 2001.

Skloot, Rebecca. *The Immortal Life of Henrietta Lacks*. New York: Random House, 2010.

Tyson, Neil DeGrasse. *Death by Black Hole and Other Cosmic Quandaries*. New York: W. W. Norton & Company, 2007.

Watson, James D. *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*. New York: Simon & Schuster, 1968.

Zimmer, Carl. *A Planet of Viruses*. Chicago: University of Chicago Press, 2012.

———. *Evolution: The Triumph of an Idea*. New York: Harper Perennial, 2006.